# Credibility and backlash*

David Foster†

January 21, 2021

**Abstract**

Recent politics has been characterized by politicians' harsh anti-immigration appeals and backlash against immigrants. I present a novel explanation for this backlash that hinges on politicians' ability to make such appeals credible. The starting point is a cheap talk model in which a politician (sender) is aligned with one of two opposed groups (receivers) and seeks to communicate her preferences to win support. Importantly, an increase in the weaker group's capacity may enable credible communication by the opposed type of politician, ironically making the weaker group worse-off. Illustrating the model, I discuss how Donald Trump credibly communicated alignment with anti-immigration groups in 2016 through harsh messaging against immigrants, whose power was increasing. More broadly, the model and case show how the behavior of strategic actors can underpin realignments, with shifts in relative group power proving crucial in enabling politicians to assemble novel political coalitions.

**Keywords:** formal model; immigration; backlash; cheap talk; campaign credibility

*For helpful comments, I thank Sean Gailmard, Jieun Kim, Anna Mikkelborg, Ben Ogden, Eric Schickler, Alexander Stephenson, Ian Turner, Rob Van Houweling, Joseph Warren, Alan Yan, and seminar participants at Berkeley, the 2020 Meeting of the Southern Political Science Association, and the 2020 Emory Conference on Institutions and Lawmaking. All errors are of course my own.

†University of California, Berkeley (foster@berkeley.edu).

In the standard narrative of political backlash, a disadvantaged social group makes incremental gains, winning small increases in rights and power. But an advantaged group subsequently feels threatened by its relative loss of status. Motivated by a perception of threat, the advantaged group fights to reverse the gains made by its opponent. Where the disadvantaged group achieves "two steps forward," the advantaged group fights to push it "one step back" (Klarman 1994; Alter and Zürn 2019). This story has intuitive plausibility, and it seemingly has recurred numerous times in American political history (Klinkner and Smith 2002, 324).

Yet a key piece is missing. Most immediately, where does the sense of threat come from? One story about backlash against immigration—the present substantive focus—holds that voters' personal exposure to increasing numbers of immigrants creates a sense of threat. Some work seems to support this hypothesis (Hawley 2011; Enos 2014, 2016; Mayda, Peri, and Steingress 2018). But other work shows no relationship between exposure to minorities and attitudes and behavior. Though Abrajano and Hajnal (2015, ch.4) find an effect of state-level Latino population on political views, they fail to find an effect of zip code-level Latino population. Jardina (2019, 97-9) suggests a weak relationship between geography and white identity. Reny, Collingwood, and Valenzuela (2019) use survey data to show no effect of an increase in the Latino population on shifts to Republicans from 2012 to 2016. Finally, Hill, Hopkins, and Huber (2019) use precinct-level data to examine the effect of changing demographics on pro-Republican shifts in voting patterns from 2012 to 2016. They find that an increase in the Hispanic population led to *less* support for the Republican candidate, as did an increase in the non-citizen foreign-born population. At best then, the evidence for local demographics leading to a backlash is mixed, with Hill, Hopkins, and Huber suggesting that the immigration issue may be nationalized.

To the extent that voters are not reacting to local demographic shifts, they likely rely on political elites to shape their perception of demographic change or sense of threat. Recent work convincingly argues that elites play a key role in shaping voters' views (Lenz 2012;

Flynn, Nyhan, and Reifler 2017). Indeed, the literature on backlash emphasizes the importance of elites and institutions in making race or immigration a political issue (Frymer 2005; Weaver 2007; Abrajano and Hajnal 2015, 35), with experimental work suggesting that elites may do so by raising the prospect of increasing diversity (Outten et al. 2012; Craig and Richeson 2014a, 2014b; Danbold and Huo 2015). These elites may include not only politicians but also the media. Exerting great influence over voters (Gilliam and Iyengar 2000; Kellstedt 2000), media companies' motivations increasingly reflect partisan political priorities (Levendusky 2013; Hedding et al. 2019). Yet whether politicians or media, when strategic actors are key to exploiting voters' potential for backlash, the standard story now exhibits an inconsistency. If such an actor knew that an opponent was about to achieve policy victories, why not activate sympathetic voters *before* the opponent succeeded and became at least partially entrenched? In other words, why not zero steps forward?

I argue that a key challenge faced by politicians seeking to stop opposed groups' victories is credibly communicating their alignment with allied groups. I first present a baseline model in which a politician sends a public cheap talk message that communicates alignment with one group by communicating opposition to the other. Concretely, when a presidential candidate expresses concerns about Medicare for All, she may communicate alignment with insurance companies, specifically because doing so alienates more radical reformers. Or when a governor issues an order directing a committee to study bathrooms, for example, he may communicate that his top priority is social conservatism, specifically because doing so alienates business interests. Following this, groups may decide whether to offer support to the politician. Finally, the politician uses this support to implement policy.

Importantly, the presence of the second group allows for mutual discipline and enables credible communication from the politician to both groups. In plain language, the politician can earn one group's support by repudiating that of the other group. But for this to be credible, we need each type of politician to benefit most from the support of the aligned group (the *single-crossing property* in this setting); otherwise, all politician types would always express alignment with the higher-capacity group. This may hold when two conditions coincide. First, a group's support and the specific goals of the type of politician with which it is aligned are complementary. Second, the two groups are relatively close in their capacity to provide resources to the politician.

This second condition underlies the connection between shifts in power and backlash. If one group is much weaker, all politicians want to express alignment with the stronger group irrespective of the truth. But should the capacity of the weaker group increase moderately, politicians may separate, with the stronger group thus able to identify its allies to promote policies that hurt the weaker group. Ironically then, the weaker group's increase in capacity may actually make it worse off. This result may be surprising, but it constitutes a fully strategic explanation for important aspects of backlash politics. To the extent that strategic elites generate voters' backlash, this theory provides a clear resolution to the initial puzzle while incorporating the central role of political communication.

Normatively, there is some cause for pessimism. Weaker groups may face backlash should their power grow. And when groups are allowed to invest in capacity, the threat posed by backlash may stop a weaker group from doing so even when capacity is free. More broadly, though, one may interpret this paper's argument as a challenge to fatalism about immigration's political consequences. Such fatalism is exemplified by Hillary Clinton, who said of European immigration, "[T]hat is what lit the flame.... Europe has done its part, and must send a very clear message—'we are not going to be able to continue provide refuge and support'—because if we don't deal with the migration issue it will continue to roil the body politic" (Wintour 2018). Similarly, in his book *Melting Pot or Civil War?*, Reihan Salam (2018) argues that if the immigration system is not reformed to emphasize high-skilled immigration and promote assimilation, the presence of low-skilled immigrants necessarily leads to racial polarization. This paper's argument suggests that negative consequences may be caused not by immigrants (unavoidably) interacting with their neighbors, but rather by particular strategic behavior of elites.

I illustrate the theory with a case study of US immigration politics. Following the enactment of the Immigration and Nationality Act of 1965, Republican politicians almost uniformly promised increased enforcement but also gestured toward sympathy for immigrants. In 1986, President Reagan signed the Immigration Reform and Control Act, which was to increase enforcement of immigration laws, yet the number of undocumented immigrants later spiked. Elite immigration hardliners blamed insufficient commitment by Reagan, George H.W. Bush, and others, and they expressed skepticism of the motives of subsequent Republicans pursuing other reform efforts. Years later, the picture changed. The country's continued diversification allowed immigrant groups to organize more effectively, enabling 2016 presidential candidate Donald Trump's strident repudiation of them to be meaningful. This sent a credible message to elite immigration foes, whose support helped propel him to the White House to pursue draconian immigration policies. Immigrants' increasing strength thus enabled credible messaging and mobilization against them.

## Prior work

The most closely related theoretical article is Farrell and Gibbons (1989), who study a cheap talk model with one sender and two receivers. Equilibrium play in my baseline model yields a structure of payoffs that can be mapped to their case of $v_1 < 0$ and $v_2 > 0$, with $w_2$ negative. As depicted in their Figure 2, either mutual discipline or no communication may result. Other work similarly explores this logic in distinct settings. In an example pertaining to campaign credibility, Harrington (1992) features a set of voters and two candidates, with all three holding private information about their own preferences. Each candidate values policy and holding office, and holding office is worth more when the voters agree with the candidate's policies. This enables candidates to separate and credibly communicate their policy intentions.[1] In an example pertaining to campaign finance, Schnakenberg and Turner (2021) examine whether a campaign contribution can signal private information about policy to a politician through its effect on the probability that the politician is re-elected. When a donor gives to a potentially misaligned moderate, it signals to the moderate that the donor has learned that light regulation of its industry is socially optimal; this is made credible by forgoing the opportunity to support the electoral prospects of the moderate's opponent, who is an ideological ally of the donor.

In contrast, this paper examines where the payoff structures that imply either mutual discipline or no communication ultimately come from as they specifically pertain to politicians' communication to groups offering support. Studying this setting leads to novel theoretical insights. To give one example, I consider an extension to the baseline model in which groups may choose to invest in capacity before playing the moves of the baseline game. Essentially, before playing a two-receiver cheap talk game together, the two receivers strategically interact to determine the inputs into the cheap talk game. The closest theoretical analogue in the literature is Antic and Persico (2020), though they study endogenous conflict of interest between a single sender and receiver under the canonical preference structure of Crawford and Sobel (1982).[2] I show that one receiver may decline free capacity in order to prevent the other's credible communication. This is reminiscent of some results on credible delegation (Gailmard and Patty 2012, 368, 374), particularly if there is a sense of capacity to review and revise the decisions of an agent. Of course, this is distinct from capacity's present role in conditioning credible communication.

This paper's substantive contribution is to show how backlash against shifts in group

1. Harrington (1993) extends this argument to a repeated setting in which players have heterogeneous beliefs about the most effective policy.

2. Related work endogenizes information acquisition (Austen-Smith 1994; Argenziano, Severinov, and Squintani 2016; Deimen and Szalay 2019), which is distinct from what is explored presently.

power may be rooted in the strategic behavior of political elites, with elite communication playing a key role. This contrasts with some existing work on backlash in American politics, which is centered around lay people's myopia (Ura 2014) and direct observation of diversification (Abrajano and Hajnal 2015).[3] However, my theory complements and extends work that has seen backlash as a product of the actions of the media (Gilliam and Iyengar 2000; Kellstedt 2000) and political elites (Weaver 2007).[4]

## The model

I present a model in which a politician communicates her preferences to two groups. Following this, the groups can grant support to help the politician implement policy. Two key aspects of the model can make the politician's communication credible. First, there is an aligned group as well as an opposed group. When the official's message signifies alignment with one group, it simultaneously signifies disagreement with the other group. Second, a group's effort to help the politician implement policy is more effective when the politician agrees with the group. Otherwise, the politician would always want to express alignment with the group facing an arbitrarily lower cost of effort, regardless of actual alignment.

3. Studying a slum neighborhood in Uganda, Habyarimana et al. (2007) relatedly argue that ethnic diversity undermines public goods provision specifically because of how lay coethnic and non-coethinic individuals interact. The present theory may alternatively suggest a role for elites.

4. Other theoretical work seeks to explain populist backlashes against economic shifts, with voters specifically reacting either to international trade (Grossman and Helpman 2018; Karakas and Mitra 2020), potential corruption by politicians and elites (Acemoglu, Egorov, and Sonin 2013), or both (Pastor and Veronesi 2018). By contrast, this paper is concerned with conflict between social groups. Additionally, unlike this paper, much of this literature assumes preferences that are nonstandard in various ways. For example, voters in Pastor and Veronesi derive utility not only from consumption but also from low inequality itself, and Grossman and Helpman incorporate social-psychological considerations.

## Formal definition

### Preliminaries

A policy $x$ lies in a policy space $\mathbb{R}$. Players consist of a politician $P$ and two groups $A$ and $B$. Policy is initially be located at $x = 0$. $P$ has one of two types corresponding to sharing preferences with either $A$ or $B$. $P$ first sends a cheap talk message. Next, groups $A$ and $B$ can grant nonnegative support to $P$ to enable $P$ to move policy. An exogenous fraction $\phi \in [0, 1)$ of each level of support must either be used to move policy in the direction preferred by the group or disposed, while the remaining fraction $1 - \phi$ may be used however $P$ prefers. The distance that $P$ may move policy is be equal to the amount of support available and usable for a given direction.

### Utility functions

Players shall have the following utility functions:

$$\begin{aligned}
U_P(x) &= \sigma x, \\
U_A(x) &= -x - \frac{{s_A}^2}{2\psi_A}, \\
U_B(x) &= x - \frac{{s_B}^2}{2\psi_B},
\end{aligned}$$

where $\sigma \in \{-1, 1\}$ is $P$'s type, $s_I$ is Group $I$'s level of support for $P$, and $\psi_I$ is Group $I$'s "capacity" or inverse marginal cost of granting support. Notice that $P$ does not have direct utility over support; $P$'s concern for it follows from its necessity to shift policy.

### Sequence of moves

The sequence of moves is as follows:

1. $P$'s type $\sigma \in \{-1, 1\}$ is drawn and revealed to $P$. With probability $p \in (0, 1)$, $\sigma = -1$

and $P$ agrees with $A$. Otherwise, $P$ agrees with $B$.

2. $P$ sends a message $m \in \{L, R\}$.

3. Each group $I \in \{A, B\}$ chooses a level of support $s_I \in \mathbb{R}^+$.

4. $P$ selects policy $x$ subject to $x \in \big[-\big(s_A + (1-\phi)s_B\big), \big((1-\phi)s_A + s_B\big)\big]$.

5. The game ends and payoffs are realized.

### Assumptions

The following assumption is without loss of generality:

**Assumption 1** (Relative Group capacity)**.** $\psi_A \leq \psi_B$.

That is to say, Group $A$ faces a higher cost of granting support.

### Summary

The exogenous parameters are $p$, $\phi$, $\psi_A$, and $\psi_B$. The endogenous choices are $m$, $s_A$, $s_B$, and $x$. The random variable is $\sigma$. As a sequential game of imperfect information, the natural equilibrium concept is perfect Bayesian equilibrium (PBE).

## Discussion

The message can represent a politician taking a symbolic action, such as the president issuing a substantively meaningless executive order, or it can represent a politician's campaign communications, such as an expression of support for a policy priority. Subsequent real-world political support (in its various forms) corresponds to the level of support in the model, and a real-world politician later issuing consequential executive orders or pushing for consequential legislation corresponds to policy implementation in the model.

A key assumption is that support from a group exhibits complementarity with the goals of the politician type who is aligned with that group; similar assumptions appear in related work (Harrington 1992, 1993; Schnakenberg 2014).[5] The degree of complementarity is represented by the parameter $\phi$. If political support took the form of money or one's own individual vote, $\phi$ would equal zero, as money and votes can be immediately and perfectly repurposed for whatever end is desired. Yet this is often not the form that it takes. Achieving policy goals can require mobilizing outside forces such as activists, interest groups, and lay people (Andrews 2001; Edwards III 2009; Bueno de Mesquita 2010). These outside forces may be better-equipped to achieve policies that they support, as achieving specific goals can mean being embedded in the right policy, donor, or activist network (Plehwe 2014; Skocpol and Hertel-Fernandez 2016; Hertel-Fernandez, Skocpol, and Sclar 2018) as well as understanding how to talk to and motivate would-be allies (Lilleker 2006, 186). If these groups' goals are actually not aligned with those of the politician, their efforts to help the politician achieve her preferred policy are likely to be ineffective. For example, it may be futile for a politician to misrepresent as an immigration opponent and subsequently call on immigration opponents to push for more lax immigration laws.[6]

## Analysis

I first examine how $A$ and $B$ should support $P$ as a function of their posterior belief about the probability that $\sigma = -1$, which I denote $\mu$. Expected utility to $A$ as a function of $s_A$ is

5. Relatedly, Ting (2011) and Hirsch and Shotts (2012) study the ability of a bureaucrat, legislature, or committee to learn "policy-specific" information, which can only be used to implement a specific policy.

6. An alternative interpretation of $\phi$ is as a reduced-form reputational cost of misrepresentation (Schnakenberg and Turner 2019, 770). Immigration opponents may refuse to carry out pro-immigration commands, and immigration supporters may not reemerge either, doubting that someone who was actually committed to their cause would ever have expressed opposition to it.

as follows:

$$\mathbb{E}U_A(s_A) = \mu\big(s_A + (1-\phi)s_B\big) + (1-\mu)\big(-(1-\phi)s_A - s_B\big) - \frac{{s_A}^2}{2\psi_A}.$$

In mirror image, the following holds for $B$:

$$\mathbb{E}U_B(s_B) = \mu\big(-s_A - (1-\phi)s_B\big) + (1-\mu)\big((1-\phi)s_A + s_B\big) - \frac{{s_B}^2}{2\psi_B}.$$

As we see, support helps $P$ to move policy. But if $P$ is the "wrong" type, she cannot perfectly repurpose support, as reflected by $\phi < 1$. The respective first-order conditions imply the following optima (with second-order conditions satisfied):

$$s_A^*(\mu) = \max\big\{\big(-(1-\phi) + \mu(2-\phi)\big)\psi_A, 0\big\},$$
$$s_B^*(\mu) = \max\big\{\big(1-\mu(2-\phi)\big)\psi_B, 0\big\}.$$

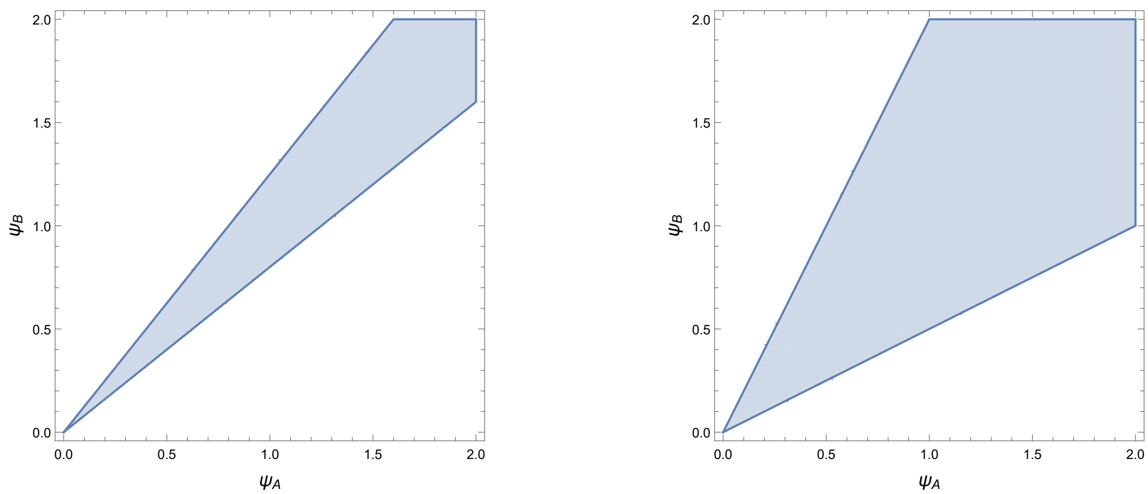
As $\mu$ increases, $A$ becomes more willing to support, because $P$ is more likely to be aligned, and likewise for $B$ given a decrease in $\mu$. Of course, both $A$ and $B$ are willing to support more when $\phi$ increases, as their support becomes more specific to their objectives and only helps move policy in their respective preferred directions.

We may now analyze the equilibria. As with a canonical cheap talk game, a pooling equilibrium always exists. The first proposition summarizes when separation is possible:

**Proposition 1** (Separation)**.** *A separating equilibrium exists whenever* $1-\phi \leq \frac{\psi_A}{\psi_B} \leq \frac{1}{1-\phi}$.

*Proof.* All proofs are in Appendix D. □

To gain intuition, this condition can be re-expressed as the intersection of two conditions: $(1-\phi)\psi_B \leq \psi_A$ and $(1-\phi)\psi_A \leq \psi_B$. That is to say, the amount that $A$ wants to offer an




(a) $\phi = 1/5$ (b) $\phi = 1/2$

Figure 1: The region in $(\psi_A, \psi_B)$-space in which separation is possible.

aligned type of $P$ must exceed the amount that $P$ could gain by misrepresenting herself as aligned with $B$, and the other way around. This allows separation to occur. See Figure 1.

Two parameter shifts that can bring the separating equilibrium into existence are of interest. First, increasing $\phi$ helps both of these conditions to be satisfied. Intuitively, the less that $P$ can use support for purposes contrary to the intentions of the groups, the less incentive $P$ has to misrepresent and take help from an opponent. Second, making $\psi_A$ and $\psi_B$ sufficiently close also helps satisfy the conditions. Intuitively, when the two groups have close to equal capacity, $P$ no longer has an incentive to communicate that she is aligned with a group solely because it has higher capacity.

## Equilibrium selection

Farrell and Gibbons (1989, 1220) demonstrate that whenever a separating equilibrium exists in a cheap talk game with one sender and two receivers, the pooling equilibrium fails the criterion of *neologism-proofness* as long as the receivers' mappings from beliefs to actions satisfy a type of consistency with one another (*coherence*). The idea behind neologism-proofness is that the sender and receivers have access to a rich language with common and literal meaning. Essentially, the pooling equilibrium is selected against because of the idea that the sender would be able to make a speech like "I really am of type 1, and you should believe me because only a type 1 sender would have an incentive to convince you so" (Farrell 1993). While coherence is defined in a setting in which receivers have binary actions, its purpose is to ensure that the sender prefers separation. This holds presently:

**Lemma 1** (Politician preference for equilibrium). *When the separating equilibrium exists, the politician prefers it to the pooling equilibrium.*

Under pooling, both groups may grant support when it is hard enough to repurpose and when their prior belief that the politician is aligned is sufficiently great. But the inability to identify friends and enemies leaves this a speculative exercise, reducing the total amount that the politician receives in aggregate as well as the amount that the politician can use to achieve preferred objectives. For this reason, each politician type does better when she can credibly identify herself to both groups.[7] Given this, I reach the following result:

**Proposition 2** (Equilibrium selection). *When the separating equilibrium exists, the pooling equilibrium fails neologism-proofness.*

Consequently, I shall select the separating equilibrium when it exists.[8]

7. This is distinct from a main result of this paper, which is that separation is not necessarily better for a group. This stems from an asymmetry: for the politician, separation assures her of finding an ally. But for a group, separation might only find its opponent an ally. The inclusion of multiple politicians with independently-drawn types would not change this, as it is not clear that they would interact in any way, and whether any given politician separated would be independent of that same question for any other politician.

8. Alternatively, Harrington (1992, 265-7) adapts the equilibrium refinement of *announcement-proofness* (Matthews, Okuno-Fujiwara, and Postlewaite 1991) to a setting with multiple senders and receivers. It is straightforward to demonstrate that this refinement also selects the separating equilibrium presently.

## Capacity shifts and backlash

We can now look at how an exogenous shift in group I's capacity, $\psi_I$, affects policy outcomes. Importantly, this may have two different effects.

First, holding fixed whether $A$ and $B$ have been able to learn the type of $P$, a group's increase in capacity straightforwardly gives it greater ability to provide support when it deems doing so to be helpful. Consider two different *informational baselines*: one in which communication is prohibited, and another of perfect information. Outcomes under the former correspond to those in the pooling equilibrium, while outcomes under the latter correspond to those in the separating equilibrium when supportable. I reach the following conclusion:

**Proposition 3.** *Within each informational baseline, expected policy* $\mathbb{E}[x]$ *weakly decreases with an increase in* $\psi_A$*, with the decrease strict whenever* $s_A^* > 0$.

That is to say, if $A$'s capacity does not determine what $P$ is able to learn, then increasing that capacity always causes policy to move in $A$'s preferred direction.

But second, a group's increase in capacity may change whether $A$ and $B$ are able to learn $P$'s type in the first place. Recalling that $A$ is the disadvantaged group, I shall examine how expected policy behaves around the value of $\psi_A$ at which separation becomes possible. Specifically, recall that $A$'s capacity has increased enough to admit separation when $\psi_A = (1-\phi)\psi_B$. At the instant that $\psi_A$ reaches this level, what happens to expected policy? Before answering this question, I establish some definitions:

**Definition 1.** *If* $p > \frac{1-\phi(1-\phi)}{2-\phi(3-\phi)}$*, then* ***separation strongly favors A****.*

**Definition 2.** *If* $p < \frac{1-2\phi}{2-\phi(3-\phi)}$*, then* ***separation strongly favors B****.*

Figure 2 illustrates where in the parameter space each of these conditions is satisfied. Roughly, when separation strongly favors $A$, $p$ is large and $\phi$ is small. And when sepa-

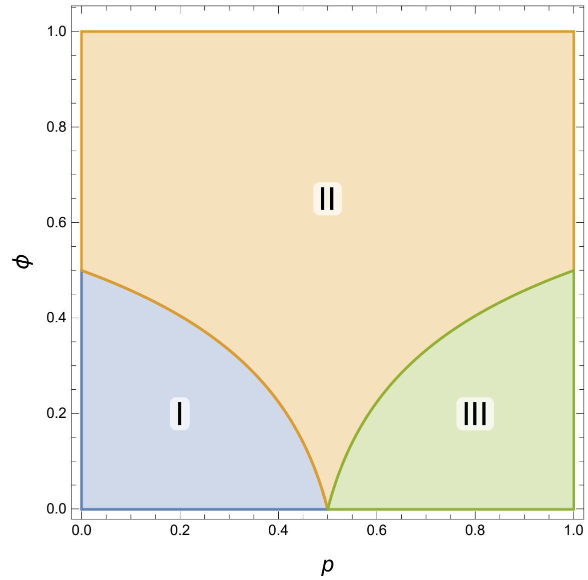



Figure 2: In region I, separation strongly favors $B$. In region II, separation strongly favors neither. In region III, separation strongly favors $A$.

ration strongly favors $B$, $p$ and $\phi$ are both small. I now reach the following result:

**Proposition 4.** *Suppose that separation does not strongly favor $A$. As a function of $\psi_A$, expected policy $\mathbb{E}[x]$ exhibits a positive jump discontinuity at $\psi_A = (1 - \phi)\psi_B$.*

When separation does not strongly favor $A$ (as in most of $(p, \phi)$-space), $A$ experiences a *backlash jump* in expected policy; this is illustrated in Figure 3.[9] Policy's sharp rightward jump at $\psi_A = (1 - \phi)\psi_B$ hurts $A$'s policy goals. This happens because when separation becomes possible due to the increase in $\psi_A$, $A$ is nevertheless still weaker than $B$. Although both are now able to identify when the politician is an ally, $B$'s still greater capacity allows it to take better advantage of this information. Only if $A$'s capacity sufficiently increases beyond $(1 - \phi)\psi_B$ is $A$ actually better off.

9. If separation *had* strongly favored $A$, we would have concluded that $p$ is large and $\phi$ is small. Under such case, policy actually exhibits a *negative* jump (benefiting $A$) at $\psi_A = (1 - \phi)\psi_B$. That is because it is difficult to grant support that only the aligned type of politician can use, but the probability that any given politician agrees with $A$ is high.

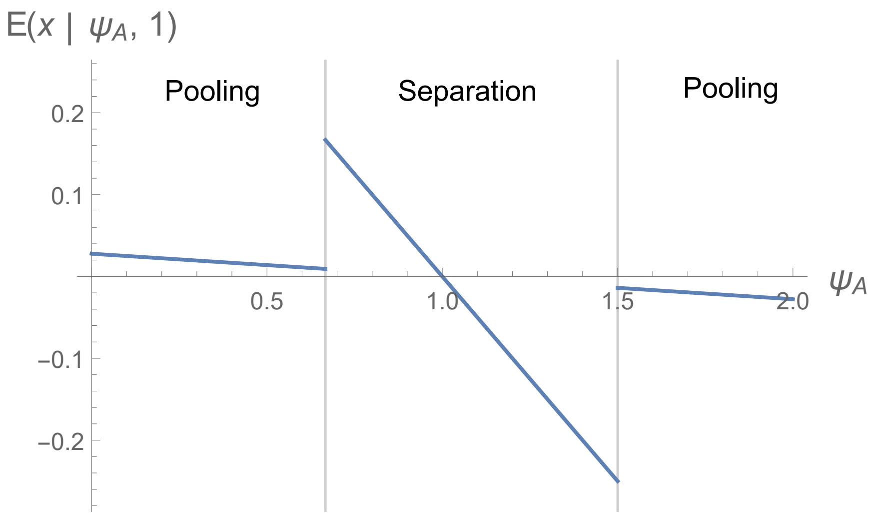



Figure 3: Expected policy as a function of $A$'s capacity $\psi_A$, with $p = 1/2$, $\phi = 1/3$ and $\psi_B = 1$. Starting from $\psi_A < 2/3$, increasing $\psi_A$ to $2/3$ brings the separating equilibrium into existence. This allows the type of $P$ that agrees with $B$ to identify herself, motivating $B$'s support and shifting expected policy rightward against $A$'s interests. Increasing $A$'s capacity only benefits it when the increase is sufficiently large.

In fact, when separation strongly favors $B$, the only way that $A$ is able to get policy back to where it was before separation became possible is for its capacity to increase so much that pooling occurs due to $B$ being comparatively low-capacity:

**Proposition 5.** *Suppose that separation strongly favors $B$. Then*

$$\lim_{\psi_A \uparrow (1-\phi)\psi_B} \mathbb{E}[x] < \lim_{\psi_A \uparrow \frac{1}{1-\phi}\psi_B} \mathbb{E}[x].$$

In plain language, the best policy for $A$ under pooling (when separation is impossible because of $A$'s low capacity) is better than the best policy for $A$ under separation (when admitted). The backlash jump is not rectified until $A$'s capacity $\psi_A$ increases beyond $\frac{1}{1-\phi}\psi_B$. This happens because separation strongly favoring $B$ means that $p$ and $\phi$ are small. Then it is difficult to grant support that only the aligned type of politician can use, but the probability of agreement with $B$ is high. Separation therefore has a large negative effect on $A$'s utility.

I conclude that unless separation strongly favors a weaker group, increased capacity can actually hurt it. The presence of a sufficiently strong opponent, and consequently the opportunity to repudiate its support in a meaningful way, enables allies of the still stronger group to credibly identify themselves. This motivates the stronger group to support the allied politician, who uses it to undermine the weaker group's goals. Thus, strengthening the weaker group can cause a policy shift against its preferences, constituting a backlash.

# Extensions

Extending the baseline model yields additional insights into credibility's role in backlash politics. I summarize the most important results here, with formal analyses in the appendices.

## Endogenous capacity ($\psi_I$)

So far, I have assumed that $\psi_A$ and $\psi_B$ are exogenous. Yet arguably, groups have the ability to invest in capacity. Given the results I have reached so far, how might this investment actually play out? In an extension, I investigate this question by supposing the existence of a group that initially has relatively low capacity (call it $A$) and another that has relatively high capacity (call it $B$). The lower-capacity group can choose to invest in capacity, followed by the ability of the higher-capacity group to respond with its own investment. Subsequently, the baseline model plays out as before. I show that in most of the parameter space, the prospect of backlash leads the disadvantaged group to forgo a free increase in capacity. This is because while increased capacity may allow $A$ to find and help its friends, this allows $B$ to increase its own capacity more than it otherwise would have while still preserving separation. As a consequence, an even higher-capacity $B$ is also able to find and help *its* friends. See Appendix A for full details.

## Endogenous complementarity ($\phi$)

We may be interested in $\phi$ being selected either by $P$ or by the groups. Substantively, this may correspond to a player's choice between building different sorts of campaign infrastructure, emphasizing either donations (low $\phi$) or activist organizing (high $\phi$). This question relates to work on the nature of the relationship between groups and parties, with groups supplying not only money but also services and expertise (Skinner 2007).

### Each group $I$ chooses $\phi_I$

Let each group $I$ have its own complementarity of support, $\phi_I$. While it may seem plausible that each $I$ would want $\phi_I$ to be as large as possible, this ignores strategic interactions. When the prior probability of a politician type aligned with a lower-capacity group $A$ is sufficiently high, a higher-capacity group $B$ may choose $\phi_B$ sufficiently small so as to jam the ability of $A$ to identify friends, since those friends would now be tempted to communicate allegiance to $B$. Remarkably, $B$'s equilibrium support is then zero, as only pooling is possible and its prior belief is that the politician is unlikely to be a friend. This therefore provides an alternative theoretical account of the "missing money" phenomenon, in which, given the enormous financial stakes of public policy, the aggregate amount of campaign donations appears smaller than it should (Chamon and Kaplan 2013). It also suggests that a stronger group may specialize in granting funds, while a weaker group may specialize in activism. See Appendix B for full details.

### $P$ chooses $\phi$

Suppose that before the baseline model plays out, the politician can determine the value of $\phi$, with a value admitting separation feasible. To rule out a trivial and implausible case, assume that the choice of $\phi$ is observable. We then have a multi-stage signaling game, to which I apply the *never dissuaded once convinced* refinement (Osborne and Rubenstein 1990, 96-8). I find that one politician type must strictly prefer separation.[10] That politician type may select a corresponding value of $\phi$. Then the other type can either select a different value of $\phi$, separating immediately, or the same value of $\phi$, only deferring separation until later. Thus, separation always occurs; see Appendix C for full details. However, there may still be a role for increasing a weaker group's capacity in enabling backlash: the minimum value of $\phi$ admitting separation in the baseline model is a decreasing function of $\psi_A$ when $\psi_A < \psi_B$.

## Polarization

One way of examining the role of polarization would be to specify two ideal points, one for each politician-group type pair. The farther apart these ideal points are, the more the environment is polarized. Then of course the position of the status quo becomes relevant. If the status quo lies sufficiently external to both ideal points, there is no longer any conflict and therefore no benefit to sending informative messages. Both groups would want to select maximum support knowing that the status quo is assured to move closer to them. And obviously there would be little sense of backlash. Greater polarization means that this situation occurs less often. The effect of increasing polarization, then, may be to increase informative messages, decrease the degree to which policy moves, and increase the possibility that a group's shift in power may lead to backlash.

What about the case in which policy lies in-between the two ideal points? If the status quo were interior but sufficiently close to one of them, the group whose ideal point was far away could only benefit from granting a lot of support. If the aligned type has arisen, policy can move a far distance favorably, while the misaligned type's potential to inflict damage would be limited. This would be reversed for the other group. So one group would want to support a lot, and the other would want to support very little. And consequently, all politician types would want to communicate alignment with the former, preventing separation from being

10. This is distinct from the result of Lemma 1, which held $\phi$ fixed.

possible. However, in the specific case in which the status quo is close to the midpoint of the ideal points *and* groups have disparate levels of capacity, sufficiently strict bounds on how far support may move policy may bring each group's effective support close to equality and enable separation when not previously possible. Of course, the finite distance between ideal points would be an upper-bound on how far policy could actually move.

In summary, while there are some ambiguities, greater polarization mostly implies more credible communication. And while in some cases this may have led to greater policy shifts, we must remember that increasing polarization decreases the measure of policies over which everyone would have agreed such that credible communication was not even necessary; in such a case, both groups would have granted support to help move policy. Therefore, for the most part, greater polarization may imply more backlash.

## The case of backlash against immigration in the US

I now illustrate the model with a case pertaining to immigration policy. To summarize, elite immigration foes long mistrusted Republican politicians' commitment to the anti-immigration cause, with politicians' communications about their preferences uninformative. But due to a recent increase in pro-immigration groups' capacity, their support became increasingly consequential. Now, a politician would be able to show alignment with anti-immigration groups by repudiating the support of pro-immigration groups. Donald Trump did exactly this with his harsh messaging, which won over elite immigration foes. This helped Trump win the election and ultimately led to a policy backlash against immigrants.[11]

11. Admittedly, some recent Republican candidates preceding Donald Trump have been unquestionably opposed to immigration, such as Tom Tancredo and Pat Buchanan. However, even if elite immigration foes of the past may have been convinced of their alignment, immigration foes faced a steeper task in elevating these less visible candidates in the absence of a diversity of social media, fund-raising platforms, and partisan news organizations outside the control of the establishment (Steger 2016; Greenfield 2016). These candidates'

## Pooling equilibrium

For years, Republican politicians promised increased enforcement but also gestured toward sympathy for Mexican migrants. For example, in a 1980 primary debate between George H.W. Bush and Ronald Reagan, Bush stated, "[A]s we have made illegal some types of labor that I would like to see legal, we're doing two things. We're creating a whole society of really honorable, decent, family-loving people that are in violation of the law, and second we're exacerbating relations with Mexico. These are good people, strong people—part of my family is Mexican." The ostensibly more conservative Reagan nevertheless felt compelled to echo Bush, stating, "Rather than talking about putting up a fence, why don't we work out some recognition of our mutual problems, make it possible for them to come here legally with a work permit" (Lee 2017). This corresponds to the pooling equilibrium.

In 1986, President Reagan signed the Immigration Reform and Control Act, whose authors had "gutted the employer sanctions"; following this, Border Patrol's staff remained relatively constant until 1993 (Plumer 2013). Jerry Kammer of the anti-immigration Center for Immigration Studies believed that this was because "Reagan was never committed to the worksite regulation that was essential to the effort to control the border" (2019). The 1986 law was followed by a sharp increase in the population of undocumented immigrants, going from 3.5 million in 1990 to about 11 million in 2005 (Passel and Cohn 2019). This perceived failure led hardliners to be skeptical of subsequent attempts to reform immigration. Writing in the conservative *American Interest*, Gallagher (2016) wrote, "[T]he 2007 Comprehensive Immigration Reform Act and the 2013 Gang of Eight bill were the same basic compromise, with tweaks and a 'trust us, this time we mean it.' Only, many people don't." Conservative columnist and strident immigration opponent Ann Coulter was blunter, writing,

> The amnesty came, but the border security never did. Illegal immigration sextupled. There have been a half dozen more amnesties since then, legalizing millions

lack of viability itself may have enabled them to credibly communicate their opposition to immigration.

> more foreigners who broke our laws. Perhaps we could have trusted Washington's sincerity thirty years ago, but Americans have already been fooled once—then, six more times. They aren't stupid. (Coulter 2015a, 8)

Once again, this corresponds to pooling in the model, with politicians unable to credibly communicate their opposition to immigration.

### Moderate increase in $\psi_A$

Soon enough though, a rising proportion of Latino immigrants led to an increase in their political capacity (corresponding to an increase in $\psi_A$). This occurred through a number of causal channels. Ramírez (2013) credits the rise of a Latino voting bloc and an increase in Latino elected officials (Zepeda-Millán 2017, 38); more specifically, Zepeda-Millán (183-4) describes efforts by Latino political organizations to encourage naturalization and voter registration. Additionally, both Ramírez (30-53) and Zepeda-Millán (67-100) point to the central role of Spanish-language media, whose existence and influence depends on a critical mass of consumers, in organizing political action. And Zepeda-Millán (127-8) notes that in cities with higher foreign-born and undocumented Mexican populations, these media's calls to political action have been more effective, specifically during the 2006 immigration reform protests. Coordinated by pro-immigration groups and the Spanish-language media, millions of people protested against the anti-immigration Sensenbrenner bill, which sought to make undocumented status a felony, among other things (11).

These protests were an important milestone in pro-immigration forces' increase in capacity, and their efforts helped to defeat the bill. Yet they had only attained intermediate capacity. As the head of a DC-based pro-immigration group summarized it, "We were strong enough to collectively stop Sensenbrenner, but not strong enough to pass comprehensive immigration reform" (174). As predicted in the model, the moderate increase in capacity marked the beginning of a backlash against immigration reform efforts. Fox News took the opportunity to stir up fear of immigrants (142). And according to activists, the protests had a polarizing effect on members of Congress, with anti-immigration groups using them to raise money (172-3). Pro-immigration activists later expressed doubt about the wisdom of these protests, concluding that they had hurt their chances at achieving comprehensive immigration reform (171). This suggests support for the endogenous capacity extension's result that a weaker group might decline to invest in capacity because it anticipates a backlash.

## Separating equilibrium

As late as 2012, even conservative television personality Sean Hannity was saying that he had "evolved" on immigration and supported a pathway to citizenship for undocumented immigrants without criminal records (Weiner 2012). Yet in this new separating equilibrium, the role of politicians' credible communication in producing backlash became clear just a few years later. In 2015, Donald Trump shattered the old messaging at his campaign announcement, famously stating that "[w]hen Mexico sends its people, they're not sending their best.... They're bringing drugs. They're bringing crime. They're rapists" (Burns 2015). After the San Bernardino mass shooting that December, he called for "a total and complete shutdown of Muslims entering the United States" (Wolf 2018). The following June, Trump claimed that a federal judge presiding over lawsuits against Trump University had "an absolute conflict" because of his "Mexican heritage" (Kendall 2016). These are only a few examples.

Contemporaneous observers argued that this strategy was costing Trump the potential support of moderates (Berenson 2016). It also appeared to hurt Trump with more diverse groups: as Greenfield (2016) noted at the time, "Trump's loaded, inflammatory language about immigration, biased 'Mexican' judges, women and the African-American experience have him polling at historically low levels with minorities and women." But enraging these constituencies was precisely what helped Trump's message resonate with elite immigration opponents. According to Coulter, "When someone like Trump comes along and is actually serious about winning the very causes the GOP purportedly seeks to advance, he is seen as a disruptive force" (Coulter 2015b). Immigration hardliners, who up to this point had failed to find traction with political leaders, thus now saw in Trump a committed immigration opponent. As Coulter later wrote, "[Y]ou know [Trump] will do what no other Republican will: Go to Washington, kick ass, mock political correctness, build a wall, [and] deport illegals" (Coulter 2016a); her book *In Trump We Trust* came out soon after (Coulter 2016b).

Winning over figures like Coulter importantly allowed Trump to influence voters (Levitsky and Ziblatt 2019, 58). The far-right website Breitbart led a network of conservative news organizations in influencing the broader media agenda (Benkler et al. 2017; Faris et al. 2017). And partisan media messages can spread even to those who do not consume them directly (Druckman, Levendusky, and McLain 2018). With evidence suggesting that exposure to partisan media has a large effect on political behavior (DellaVigna and Kaplan 2007; Martin and Yurukoglu 2017), it is likely that Trump's ability to credibly communicate his preferences to prominent anti-immigration elites ultimately moved erstwhile supporters in the public.

Indeed, this campaign messaging turned out to be largely credible: Trump's election enabled draconian immigration policies, including the travel ban on many majority-Muslim countries and the policy of separating families at the Mexican border. Intermediated by elite immigration foes, the anti-immigration support that Trump earned during the campaign proved crucial in helping to shield such policies from opposition, at times enabling him to neutralize Republican critics. For example, Arizona Senator Jeff Flake wrote a New York Times editorial in August 2017 specifically criticizing Trump's immigration stances and arguing that the U.S. benefits from unskilled laborers coming from Mexico (Flake 2017). Quickly enough, supporters' response to Trump's withering criticism of "flake Jeff Flake" reinforced Flake's difficulties with the primary electorate, leading him to announce in October that he would not seek re-election (Gay Stolberg 2017).

### New coalitions

Republican priorities did not merely shift under everyone's feet. Rather, the possibility of separation allowed Republican politicians such as Trump to credibly communicate the Republican Party's commitment to opposing immigration. This may have precipitated activation of certain types of Republican voters (Sides, Tesler, and Vavreck 2018) or sorting across the parties (Cohn 2017; Sides, Tesler, and Vavreck 2017, 42). In 2002, for example, 62% of Democrats agreed in a survey that "large numbers of immigrants and refugees coming into the US" posed a "critical threat" to the country, more than the 58% of Republicans who agreed (Kafura and Hammer 2019). But by 2019, 78% of Republicans agreed while only 19% of Democrats agreed. Ultimately then, the rise of pro-immigration groups' power enabled credible messaging by Republican politicians against immigrants, thus strengthening the association between the Republican Party and restrictionism.

This case has thus demonstrated how the increasing capacity of a lower-capacity group can bring about a shift in political messaging. This messaging credibly communicates policy commitments in a way that was previously impossible, enabling political elites to construct new political coalitions that move policy against the interests of the lower-capacity group.

## Conclusion

This paper has started from the premise that a key problem for groups is credibly identifying allied politicians. One way for a politician to communicate her alignment with a group is through repudiation of an opposed group. Yet credibility requires that the two groups be close enough to one another in their capacity to offer support to an ally. When pro-immigration groups are so weak that no one would ever prefer their help over that from their opponents, neither pro- nor anti-immigration groups can believe messages from any politician. But when pro-immigration groups become stronger, repudiating them becomes meaningful. And when an opposed politician thus earns the support of anti-immigration groups, this can turn policy against the preferences of the pro-immigration groups, constituting a backlash.

This paper has thus emphasized the role of elites in producing backlash, particularly that against immigrants. Rather than looking to lay people's direct perception of demographic shifts, I have examined elites' role in shaping this perception. Such a perspective demands a model that satisfies two criteria. First, elite actors are strategic and forward-looking. Second, elite communication plays a key role. The model that I have presented satisfies these criteria. An anti-immigration group may anticipate that its opponents are about to achieve policy victories, but the group's ability to stop its opponents may be limited by the aligned type of politician's inability to credibly communicate her preferences. When opponents' capacity increases, communication becomes credible and policy victories may reverse.

More broadly, the model's focus on elites helps us to understand how the behavior of strategic actors can underpin realignments, with shifts in relative group power proving crucial in enabling politicians to assemble novel political coalitions. The success of Donald Trump's anti-immigration campaign was made possible by an increase in the capacity of pro-immigration groups. And following his campaign and presidential actions, the Republican Party has become inextricably associated with opposition to immigration. Future work should inquire further about how group power and political communication determine the shape that party coalitions take.

# References

Abrajano, Marisa, and Zoltan L. Hajnal. 2015. *White Backlash.* Princeton, NJ: Princeton University Press.

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin. 2013. "A political theory of populism." Publisher: MIT Press, *The Quarterly Journal of Economics* 128 (2): 771–805.

Alter, Karen J., and Michael Zürn. 2019. *Theorizing Backlash Politics.* SSRN Scholarly Paper ID 3493079. Rochester, NY: Social Science Research Network, November. Accessed November 12, 2020. https://papers.ssrn.com/abstract=3493079.

Andrews, Kenneth T. 2001. "Social Movements and Policy Implementation: The Mississippi Civil Rights Movement and the War on Poverty, 1965 to 1971." *American Sociological Review* 66 (1): 71–95.

Antic, Nemanja, and Nicola Persico. 2020. "Cheap Talk With Endogenous Conflict of Interest." *Econometrica* 88, no. 6 (November): 2663–2695.

Argenziano, Rossella, Sergei Severinov, and Francesco Squintani. 2016. "Strategic Information Acquisition and Transmission." *American Economic Journal: Microeconomics* 8, no. 3 (August): 119–155.

Austen-Smith, David. 1994. "Strategic Transmission of Costly Information." *Econometrica* 62 (4): 955–963.

Benkler, Yochai, Rob Faris, Hal Roberts, and Ethan Zuckerman. 2017. "Study: Breitbart-led right-wing media ecosystem altered broader media agenda." *Columbia Journalism Review* (March). Accessed October 24, 2020. https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php.

Berenson, Tessa. 2016. "Poll Shows Limits Of Donald Trump's Anti-Immigrant Rhetoric." *Time* (August). Accessed November 15, 2020. https://time.com/4464831/donald-trump-mexico-wall-election-poll/.

Bueno de Mesquita, Ethan. 2010. "Regime Change and Revolutionary Entrepreneurs." *American Political Science Review* 104, no. 3 (August): 446–466.

Burns, Alexander. 2015. "Choice Words From Donald Trump, Presidential Candidate." *The New York Times* (June). https://www.nytimes.com/politics/first-draft/2015/06/16/choice-words-from-donald-trump-presidential-candidate/.

Chamon, Marcos, and Ethan Kaplan. 2013. "The Iceberg Theory of Campaign Contributions: Political Threats and Interest Group Behavior." *American Economic Journal: Economic Policy* 5, no. 1 (February): 1–31.

Cohn, Nate. 2017. "The Obama-Trump Voters Are Real. Here's What They Think." *The New York Times* (August). Accessed November 16, 2020. https://www.nytimes.com/2017/08/15/upshot/the-obama-trump-voters-are-real-heres-what-they-think.html.

Coulter, Ann. 2015a. *Adios, America: The Left's Plan to Turn Our Country into a Third World Hellhole.* Washingtion, DC: Regnery Publishing.

———. 2016a. *G.O.P. Baffled as Voters Rally to Popular Candidate,* February. https://anncoulter.com/2016/02/17/g-o-p-baffled-as-voters-rally-to-popular-candidate/.

———. 2016b. *In Trump We Trust: E Pluribus Awesome!* New York, New York: Sentinel, August.

———. 2015b. *Useless Idiots,* September. https://anncoulter.com/2015/09/23/useless-idiots/.

Craig, Maureen A., and Jennifer A. Richeson. 2014a. "More Diverse Yet Less Tolerant? How the Increasingly Diverse Racial Landscape Affects White Americans' Racial Attitudes." *Personality and Social Psychology Bulletin* 40, no. 6 (June): 750–761.

———. 2014b. "On the Precipice of a "Majority-Minority" America: Perceived Status Threat From the Racial Demographic Shift Affects White Americans' Political Ideology." Publisher: SAGE Publications Inc, *Psychological Science* 25, no. 6 (June): 1189–1197.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic information transmission." *Econometrica* 50 (6): 1431–1451.

Danbold, Felix, and Yuen J. Huo. 2015. "No Longer "All-American"? Whites' Defensive Reactions to Their Numerical Decline." Publisher: SAGE Publications Inc, *Social Psychological and Personality Science* 6, no. 2 (March): 210–218.

Deimen, Inga, and Dezső Szalay. 2019. "Delegated Expertise, Authority, and Communication." *American Economic Review* 109, no. 4 (April): 1349–1374.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics* 122, no. 3 (August): 1187–1234.

Druckman, James N., Matthew S. Levendusky, and Audrey McLain. 2018. "No Need to Watch: How the Effects of Partisan Media Can Spread via Interpersonal Discussions." *American Journal of Political Science* 62 (1): 99–112.

Edwards III, George C. 2009. *The Strategic President: Persuasion and Opportunity in Presidential Leadership.* Princeton, NJ: Princeton University Press.

Enos, Ryan D. 2014. "Causal effect of intergroup contact on exclusionary attitudes." *Proceedings of the National Academy of Sciences* 111, no. 10 (March): 3699.

Enos, Ryan D. 2016. "What the Demolition of Public Housing Teaches Us about the Impact of Racial Threat on Political Behavior." *American Journal of Political Science* 60 (1): 123–142.

Faris, Rob, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election.* Technical report. Cambridge, MA: Berkman Klein Center for Internet & Society at Harvard University, August. `https://cyber.harvard.edu/publications/2017/08/mediacloud`.

Farrell, Joseph. 1993. "Meaning and Credibility in Cheap-Talk Games." *Games and Economic Behavior* 5, no. 4 (October): 514–531.

Farrell, Joseph, and Robert Gibbons. 1989. "Cheap talk with two audiences." *American Economic Review* 79 (5): 1214–1223.

Flake, Jeff. 2017. "We Need Immigrants With Skills. But Working Hard Is a Skill." *The New York Times* (August).

Flynn, D. J., Brendan Nyhan, and Jason Reifler. 2017. "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics." *Political Psychology* 38 (S1): 127–150.

Frymer, Paul. 2005. "Racism Revised: Courts, Labor Law, and the Institutional Construction of Racial Animus." *American Political Science Review* 99, no. 3 (August): 373–387.

Gailmard, Sean, and John W. Patty. 2012. "Formal Models of Bureaucracy." *Annual Review of Political Science* 15 (1): 353–377.

Gallagher, Nicholas M. 2016. "30 Years Since Amnesty-for-Enforcement Failure." *The American Interest* (November).

Gay Stolberg, Sheryl. 2017. "Jeff Flake, a Fierce Trump Critic, Will Not Seek Re-election for Senate." *The New York Times* (October).

Gilliam, Franklin D., Jr., and Shanto Iyengar. 2000. "Prime Suspects: The Influence of Local Television News on the Viewing Public on JSTOR." *American Journal of Political Science* 44, no. 3 (July): 560–573.

Greenfield, Jeff. 2016. "Donald Trump Is Pat Buchanan With Better Timing." *POLITICO Magazine* (September). https://www.politico.com/magazine/story/2016/09/donald-trump-pat-buchanan-republican-america-first-nativist-214221.

Grossman, Gene M, and Elhanan Helpman. 2018. *Identity Politics and Trade Policy.* NBER Working Paper 25348. December. http://www.nber.org/papers/w25348.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101, no. 4 (November): 709–725.

Harrington, Joseph E. 1993. "The Impact of Reelection Pressures on the Fulfillment of Campaign Promises." *Games and Economic Behavior* 5, no. 1 (January): 71–97.

———. 1992. "The Revelation of Information Through the Electoral Process: An Exploratory Analysis." *Economics & Politics* 4, no. 3 (November): 255–276.

Hawley, George. 2011. "Political Threat and Immigration: Party Identification, Demographic Context, and Immigration Policy Preference." *Social Science Quarterly* 92 (2): 404–422.

Hedding, Kylah J., Kaitlin C. Miller, Jesse Abdenour, and Justin C. Blankenship. 2019. "The Sinclair Effect: Comparing Ownership Influences on Bias in Local TV News Content." *Journal of Broadcasting & Electronic Media* 63, no. 3 (July): 474–493.

Hertel-Fernandez, Alexander, Theda Skocpol, and Jason Sclar. 2018. "When Political Mega-Donors Join Forces: How the Koch Network and the Democracy Alliance Influence Organized U.S. Politics on the Right and Left." *Studies in American Political Development* 32, no. 2 (October): 127–165.

Hill, Seth J., Daniel J. Hopkins, and Gregory A. Huber. 2019. "Local demographic changes and US presidential voting, 2012 to 2016." *Proceedings of the National Academy of Sciences* 116, no. 50 (December): 25023–8.

Hirsch, Alexander V., and Kenneth W. Shotts. 2012. "Policy-Specific Information and Informal Agenda Power." *American Journal of Political Science* 56 (1): 67–83.

Jardina, Ashley. 2019. *White Identity Politics.* Cambridge Studies in Public Opinion and Political Psychology. Cambridge: Cambridge University Press. Accessed October 9, 2020.

Kafura, Craig, and Bettina Hammer. 2019. *Republicans and Democrats in Different Worlds on Immigration.* Technical report. Chicago: The Chicago Council on Global Affairs, September. Accessed November 16, 2020. https://www.thechicagocouncil.org/publication/lcc/republicans-and-democrats-different-worlds-immigration.

Kammer, Jerry. 2019. *How the Washington Post Exposé of Trump Missed the Story of How Reagan Wrecked Immigration Reform,* January. https://cis.org/Kammer/How-Washington-Post-Expose-Trump-Missed-Story-How-Reagan-Wrecked-Immigration-Reform.

Karakas, Leyla D., and Devashish Mitra. 2020. "Inequality, redistribution and the rise of outsider candidates." *Games and Economic Behavior* 124 (November): 1–16.

Kellstedt, Paul M. 2000. "Media Framing and the Dynamics of Racial Policy Preferences." *American Journal of Political Science* 44 (2): 245–260.

Kendall, Brent. 2016. "Trump Says Judge's Mexican Heritage Presents 'Absolute Conflict'." *Wall Street Journal* (June).

Klarman, Michael J. 1994. "How Brown Changed Race Relations: The Backlash Thesis." *Journal of American History* 81, no. 1 (June): 81–118.

Klinkner, Philip A., and Rogers M. Smith. 2002. *The Unsteady March: The Rise and Decline of Racial Equality in America.* University of Chicago Press, April.

Lee, Timothy B. 2017. "This video of Ronald Reagan shows how much the Republican Party has changed on immigration." *Vox* (January). `https://www.vox.com/2017/1/29/14429368/reagan-bush-immigration-attitude`.

Lenz, Gabriel S. 2012. *Follow the leader?: How voters respond to politicians' policies and performance.* Chicago: University of Chicago Press.

Levendusky, Matthew. 2013. *How Partisan Media Polarize America.* Illustrated Edition. Chicago ; London: University of Chicago Press, September.

Levitsky, Steven, and Daniel Ziblatt. 2019. *How Democracies Die.* Reprint edition. New York: Crown, January.

Lilleker, Darren G. 2006. *Key Concepts in Political Communication.* London: SAGE, February.

Martin, Gregory J., and Ali Yurukoglu. 2017. "Bias in Cable News: Persuasion and Polarization." *American Economic Review* 107, no. 9 (September): 2565–2599.

Matthews, Steven A, Masahiro Okuno-Fujiwara, and Andrew Postlewaite. 1991. "Refining cheap-talk equilibria." *Journal of Economic Theory* 55, no. 2 (December): 247–273.

Mayda, Anna Maria, Giovanni Peri, and Walter Steingress. 2018. *The Political Impact of Immigration: Evidence from the United States.* NBER Working Paper 24510. April. http://www.nber.org/papers/w24510.

Osborne, Martin J., and Ariel Rubenstein. 1990. *Bargaining and Markets.* San Diego: Emerald Publishing, April.

Outten, H. Robert, Michael T. Schmitt, Daniel A. Miller, and Amber L. Garcia. 2012. "Feeling Threatened About the Future: Whites' Emotional Reactions to Anticipated Ethnic Demographic Changes." *Personality and Social Psychology Bulletin* 38, no. 1 (January): 14–25.

Passel, Jeffrey S., and D'Vera Cohn. 2019. *Mexicans decline to less than half the U.S. unauthorized immigrant population for the first time,* June. Accessed November 21, 2020. https://www.pewresearch.org/fact-tank/2019/06/12/us-unauthorized-immigrant-population-2017/.

Pastor, Lubos, and Pietro Veronesi. 2018. *Inequality Aversion, Populism, and the Backlash Against Globalization.* NBER Working Paper 24900. National Bureau of Economic Research, August. https://www.nber.org/papers/w24900.

Plehwe, Dieter. 2014. "Think tank networks and the knowledge–interest nexus: the case of climate change." *Critical Policy Studies* 8, no. 1 (January): 101–115.

Plumer, Brad. 2013. "Congress tried to fix immigration back in 1986. Why did it fail?" *The Washington Post* (January).

Ramírez, Ricardo. 2013. *Mobilizing Opportunities: The Evolving Latino Electorate and the Future of American Politics.* University of Virginia Press, November.

Reny, Tyler T., Loren Collingwood, and Ali A. Valenzuela. 2019. "Vote Switching in the 2016 Election: How Racial and Immigration Attitudes, Not Economics, Explain Shifts in White Voting." Publisher: Oxford Academic, *Public Opinion Quarterly* 83, no. 1 (May): 91–113.

Salam, Reihan. 2018. *Melting Pot or Civil War?: A Son of Immigrants Makes the Case Against Open Borders.* New York: Sentinel, September.

Schnakenberg, Keith E. 2014. "Group Identity and Symbolic Political Behavior." *Quarterly Journal of Political Science* 9, no. 2 (June): 137–167.

Schnakenberg, Keith E., and Ian R. Turner. 2021. "Helping Friends or Influencing Foes: Electoral and Policy Effects of Campaign Finance Contributions." *American Journal of Political Science* 65, no. 1 (January): 88–100.

———. 2019. "Signaling with Reform: How the Threat of Corruption Prevents Informed Policy-making." *American Political Science Review* 113, no. 3 (August): 762–777.

Sides, John, Michael Tesler, and Lynn Vavreck. 2018. "Hunting where the ducks are: activating support for Donald Trump in the 2016 Republican primary." *Journal of Elections, Public Opinion and Parties* 28, no. 2 (April): 135–156.

———. 2017. "The 2016 US election: How Trump lost and won." *Journal of Democracy* 28 (2): 34–44.

Skinner, Richard M. 2007. *More Than Money: Interest Group Action in Congressional Elections.* Rowman & Littlefield.

Skocpol, Theda, and Alexander Hertel-Fernandez. 2016. "The Koch Network and Republican Party Extremism." *Perspectives on Politics* 14, no. 3 (September): 681–699.

Steger, Wayne P. 2016. "Conditional Arbiters: The Limits of Political Party Influence in Presidential Nominations." *PS: Political Science & Politics* 49, no. 4 (October): 709–715.

Ting, Michael M. 2011. "Organizational Capacity." *The Journal of Law, Economics, and Organization* 27, no. 2 (August): 245–271.

Ura, Joseph Daniel. 2014. "Backlash and Legitimation: Macro Political Responses to Supreme Court Decisions." *American Journal of Political Science* 58 (1): 110–126.

Vincent, Daniel R. 1998. "Repeated signalling games and dynamic trading relationships." *International Economic Review* 39 (2): 275–294.

Weaver, Vesla M. 2007. "Frontlash: Race and the Development of Punitive Crime Policy." *Studies in American Political Development* 21 (2): 230–265.

Weiner, Rachel. 2012. "Sean Hannity: I've 'evolved' on immigration." *Washington Post* (November).

Wintour, Patrick. 2018. "Hillary Clinton: Europe must curb immigration to stop rightwing populists." *The Guardian* (November).

Wolf, Richard. 2018. "Travel ban lexicon: From candidate Donald Trump's campaign promises to President Trump's tweets." *USA TODAY* (April).

Zepeda-Millán, Chris. 2017. *Latino Mass Mobilization: Immigration, Racialization, and Activism.* Cambridge: Cambridge University Press.

# Supporting information for "Credibility and backlash"

January 21, 2021

# A Endogenous capacity extension

I endogenize each $\psi_I$, allowing group $I$ to choose to increase it from an initial value. Building on the results of the baseline model, I demonstrate that a weaker group may decline a free increase in capacity.

### Preliminaries

In Stage 2, the baseline model plays out as before. In Stage 1, each $I \in \{A, B\}$ starts with an initial level of capacity $\underline{\psi}_I$. At no exogenous cost, $I$ may later choose to increase $\psi_I$ up to a maximum of $\overline{\psi}_I$ (but may not decrease it).

### Sequence of moves

The sequence of moves is as in the baseline model, except preceding them is the following:

**Stage 1**

1. $A$ selects its capacity $\psi_A \in [\underline{\psi}_A, \overline{\psi}_A]$.

2. $B$ selects its capacity $\psi_B \in [\underline{\psi}_B, \overline{\psi}_B]$.

Subsequent moves shall collectively comprise Stage 2.

### Utility functions

In Stage 2, $P$, $A$, and $B$ shall have the same utility functions as before. In Stage 1, $A$ and $B$ shall have the following utility functions ($P$'s Stage 1 utility is inconsequential):

$$U_A^1(x) = -x,$$
$$U_B^1(x) = x.$$

**Assumptions**

The following assumption concerns the initial capacity of the groups:

**Assumption A.1** (Initial group capacity)**.** $\underline{\psi}_A < (1-\phi)\underline{\psi}_B$.

Corresponding to the case of interest, this simply states that $A$ starts off with lower capacity compared to $B$, such that only the pooling equilibrium is admitted.

Next, I assume the following:

**Assumption A.2** (Intermediate initial capacity for $B$)**.** $(1-\phi)\overline{\psi}_A < \underline{\psi}_B < \frac{1}{1-\phi}\overline{\psi}_A$.

The first part of this, $(1-\phi)\overline{\psi}_A < \underline{\psi}_B$, simply states that no matter $B$'s choice of investment, $A$ cannot induce pooling by becoming sufficiently higher-capacity than $B$. The second part of this, $\underline{\psi}_B < \frac{1}{1-\phi}\overline{\psi}_A$, ensures non-triviality; it would otherwise be impossible for any strategy profile to lead to separation in equilibrium.

Finally, I assume the following:

**Assumption A.3** (High potential capacity for $B$)**.** $\frac{1}{1-\phi}\overline{\psi}_A < \overline{\psi}_B$.

This simply states that no matter how much $A$ invests, $B$ can always induce pooling with sufficient investment. Results are similar without this assumption, but it greatly simplifies the analysis while corresponding substantively to the case of interest.

**Summary**

The exogenous parameters are $\underline{\psi}_A$, $\overline{\psi}_A$, $\underline{\psi}_B$, $\overline{\psi}_B$, $p$, $\phi$, $\psi_A$, and $\psi_B$. The endogenous choices are $\psi_A$, $\psi_B$, $m$, $s_A$, $s_B$, and $x$. The random variable is $\sigma$. As a sequential game of imperfect information, the natural equilibrium concept is perfect Bayesian equilibrium (PBE). I continue to apply the equilibrium selection criterion described previously.

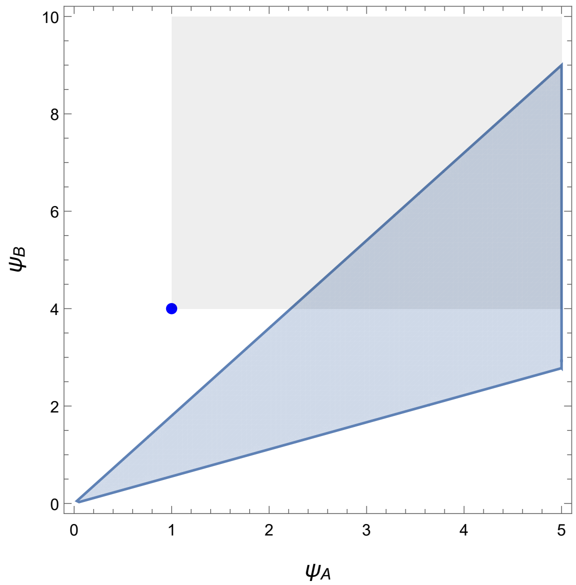



Figure A.1: An example fitting the assumptions. In particular, $\underline{\psi}_A = 1$, $\underline{\psi}_B = 4$, $\overline{\psi}_A = 5$, $\overline{\psi}_B = 10$, and $\phi = 4/9$. As before, the cone is the region in which separation occurs. The dot shows initial capacity, and the rectangle shows the set of points to which players may move capacity.

## Discussion

I comment briefly on the assumptions. First, consider the order of moves. Allowing $A$ to move first corresponds to the backlash dynamics that I explore. The question is, in anticipation of a higher-capacity group's strategic response, how does a lower-capacity group make decisions about building its capacity? The assumed order of moves fits this question.

Next, the assumption that increasing capacity is free only strengthens the results. Strikingly, we shall see that $A$ may still decline to do so.

Finally, consider the utility functions. In Stage 2, $A$ and $B$ incur a cost of supporting $P$. Yet in Stage 1, $A$ and $B$ are unconcerned with these future costs. This can be justified substantively. One can imagine the groups in Stage 1 as representing different actors compared to those in Stage 2. Donors or activists making decisions about how to build their organizations may care about policy but not about the effort that bureaucrats in the future will have to exert. Alternatively, the costs of granting support can capture a notion of constraint at the moment that it is granted rather than a source of negative utility to an institutional designer. While this assumption simplifies the analysis, it also allows us to continue to focus on the substantively interesting question of how policy actually moves.

## Analysis

Stage 2 plays out as before. In Stage 1, there are three cases. Under pooling, when $p < \frac{1-\phi}{2-\phi}$, only $B$ supports (*Case 1*), when $\frac{1-\phi}{2-\phi} < p < \frac{1}{2-\phi}$, both support (*Case 2*), and when $\frac{1}{2-\phi} < p$, only $A$ supports (*Case 3*).

A key observation is that once $A$ has made a choice of $\psi_A$, only two things can be optimal for $B$: choose $\psi_B$ just small enough such that a separating equilibrium continues to be possible, or choose $\psi_B$ as large as possible. In Cases 1 and 2, which option $B$ prefers is a function of $\psi_A$ (while in Case 3, $B$ grants zero support under pooling, so that its only consideration in selecting $\psi_B$ is which equilibrium it wishes to induce; we shall see that this is not a function of $\psi_A$). For a small value of $\psi_A$, $B$ would need to forgo a large potential increase in $\psi_B$ to maintain separation. As $\psi_A$ increases, though, this sacrifice diminishes, and setting $\psi_B = \frac{1}{1-\phi}\psi_A$ (the largest value of $\psi_A$ compatible with separation) becomes relatively more attractive. This is summarized in the following lemma:

**Lemma A.1** ($B$'s best response)**.** *Suppose that Case 1 or 2 holds. There exists a threshold value of $\psi_A$, call it $\tilde{\psi}_A$, such that $\psi_A \leq \tilde{\psi}_A$ implies that $B$ will induce pooling by setting $\psi_B = \overline{\psi}_B$, while $\psi_A > \tilde{\psi}_A$ implies that $B$ will induce separation by setting $\psi_B = \frac{1}{1-\phi}\psi_A$.*

*Suppose instead that Case 3 holds. Then $B$ either always prefers pooling or always prefers separation irrespective of $\psi_A$. If $B$ always prefers pooling, it sets $\psi_B = \overline{\psi}_B$. If $B$ always prefers separation, it sets $\psi_B = \frac{1}{1-\phi}\psi_A$.*

See Figure A.2 for an illustration of this result.[1] Effectively, when $\psi_A$ is chosen to be small,

1. To guarantee equilibrium existence, and because $A$ can move $\psi_A$ rightward or leftward from $\tilde{\psi}_A$ by any

$B$ would need to set $\psi_B$ much smaller than $\overline{\psi}_B$ to allow for separation, i.e. $\frac{1}{1-\phi}\psi_A$ is small. In such case, $B$ does better to increase capacity as much as possible and give up on separation. Yet when $\psi_A$ becomes larger, setting $\psi_B = \frac{1}{1-\phi}\psi_A$ becomes relatively more attractive, such that $B$ eventually prefers to sacrifice some capacity to allow separation to happen.

Given $B$'s best response, we shall see that $A$'s optimum can be one of two things. First, $A$ may seek to avoid separation by setting $\psi_A = \tilde{\psi}_A$. That is to say, $A$ chooses the largest $\psi_A$ compatible with pooling. Second, $A$ may select $\psi_A$ as large as possible, with either separation or pooling resulting depending on $B$'s best reply.

To help characterize equilibrium outcomes, I define cutoff values of $p$. Letting

$$T_p \equiv \frac{\overline{\psi}_A - \sqrt{t\overline{\psi}_A\left(\overline{\psi}_A - 4(1-\phi)^2\underline{\psi}_B\right)}}{2(1-\phi)^2(2-\phi)\underline{\psi}_B} + \frac{1-\phi}{2-\phi},$$

I shall say that $p$ is *low* when $p < \frac{1}{2-\phi}$, *intermediate* when $\frac{1}{2-\phi} < p < \min\left\{T_p, \frac{1-\phi(1-\phi)}{2-\phi(3-\phi)}\right\}$, *high* when $T_p < p < \frac{1-\phi(1-\phi)}{2-\phi(3-\phi)}$, and *very high* when $\frac{1-\phi(1-\phi)}{2-\phi(3-\phi)} < p$. These regions are illustrated in Figure A.3. We are now ready for the following result:

**Proposition A.1.** *When $p$ is low, $A$ sets $\psi_A = \max\{\tilde{\psi}_A, (1-\phi)\underline{\psi}_B\}$, $B$ sets $\psi_B = \overline{\psi}_B$, and pooling occurs. When $p$ is intermediate, $A$ sets $\psi_A = (1-\phi)\underline{\psi}_B$, $B$ sets $\psi_B = \overline{\psi}_B$, and pooling occurs. When $p$ is high, $A$ sets $\psi_A = \overline{\psi}_A$, $B$ sets $\psi_B = \frac{1}{1-\phi}\overline{\psi}_A$, and separation occurs. Finally, when $p$ is very high, $A$ sets $\psi_A = \overline{\psi}_A$, $B$ sets $\psi_B = \overline{\psi}_B$, and pooling occurs.*

When $p$ is low, $A$ holds back on increasing $\psi_A$ too far because it fears the consequences of separation. This is because $p$ is simply too small, such that when friends and enemies can be identified, this more often benefits the higher-capacity $B$.

Next, when $p$ is intermediate, $B$ always wants to separate: it grants zero support under

$\epsilon > 0$, I assume that $A$ can break $B$'s indifference whichever way $A$ prefers when $\psi_A = \tilde{\psi}_A$.

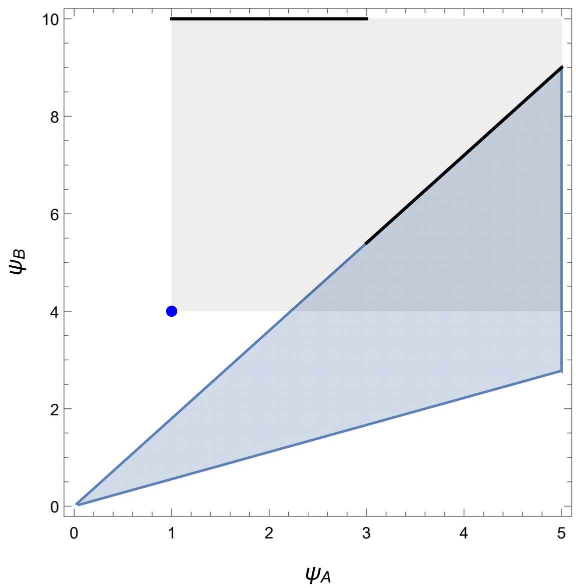



Figure A.2: Maintaining the parametric assumptions of Figure A.1 and fixing $p = \frac{207}{700}$ (so Case 1 holds), the black line is $B$'s optimal choice of $\psi_B$ given $\psi_A$. The discontinuity is at $\tilde{\psi}_A = 3$.

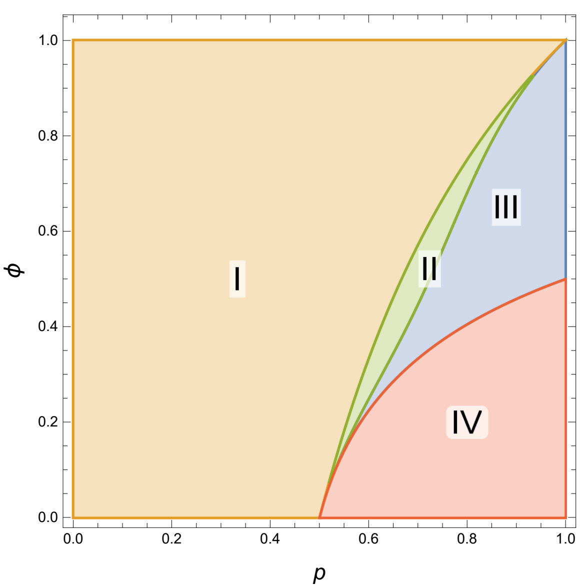



Figure A.3: In regions I, II, III, and IV, $p$ is low, intermediate, high, and very high, respectively. Very high $p$ coincides with separation strongly favoring $A$. In this example, $\overline{\psi}_A = \underline{\psi}_B = 8$.

pooling, while $p$ is tilted enough in $A$'s favor that it grants positive support. If separation were instead to occur, the higher-capacity $B$ would identify and support more friends than $A$ would like, relative to $A$'s benefit of identifying its own friends.

Next, when $p$ is high, $B$ still always wants to separate. What has changed is $A$'s calculation. Now, $p$ has become sufficiently large such that $A$'s benefit of identifying its friends improves relative to the cost of $B$ being able to identify its friends. While $B$ still does better under separation, it has become relatively attractive to $A$ compared to the alternative of keeping $\psi_A$ so small that for $B$ it is infeasible to induce separation.

Finally, when $p$ is very high, separation strongly favors $A$ in the sense defined above. Large $p$ and small $\phi$ means that most politicians are likely to be $A$'s friend. Yet without the ability to identify friends or grant support that can only be used for agreeable purposes, there is a high potential for $A$'s support to be repurposed. Therefore, $B$ always wants to induce pooling, so both players increase their capacity as far as possible.

A comparative static implication we thus see is that increasing $p$ sufficiently may make it larger than $T_p$, implying that $A$ comes to prefer separation. That is to say, when $A$ is more likely to identify a friend, it becomes more valuable for it to try to do so. Of course, increasing $p$ too much may therefore lead $B$ to induce pooling. I additionally find the following:

**Proposition A.2** (Comparative statics). *The measure of $\phi$ in which separation occurs is increasing in $\overline{\psi}_A$ and decreasing in $\underline{\psi}_B$.*

These comparative statics essentially reflect a change in various forms of relative capacity of $A$ compared to $B$. When $A$'s maximum potential capacity decreases, separation becomes less desirable to $A$. And when $B$'s initial capacity is greater, this gives $A$ room to increase its capacity more while still not triggering separation, making pooling relatively attractive. In summary, then, increasing $B$'s relative current and potential capacity leads $A$ to be increasingly wary of choosing to increase its own capacity to the maximum that is feasible.

# B Group selection of complementarity extension

I extend the baseline model to examine how groups might endogenously choose complementarity $\phi$. I therefore relax the assumption that there is a common value of $\phi$ and instead allow it to be specific to each player, i.e. $\phi_I$ is the fraction of $I$'s support that cannot be repurposed, with $I \in \{A, B\}$. Additionally, selection of each $\phi_I$ occurs simultaneously before the baseline model plays out.

## Formal definition

Preliminaries are as in the baseline model, except an endogenously chosen fraction $\phi_I$ of the support granted by group $I \in \{A, B\}$, must either be used to move policy in the specified direction or disposed. The sequence of moves is as before, except preceding them is the following:

**Stage 1**

1. Each group $I \in \{A, B\}$ simultaneously selects $\phi_I \in [\underline{\phi}_I, \overline{\phi}_I]$.

Subsequent moves shall collectively comprise Stage 2. Utility functions are as in the endogenous capacity extension. Assumption 1 is maintained. Next, to analyze a non-trivial case, I assume the following:

**Assumption B.1** (Non-triviality)**.** $\underline{\phi}_B \leq 1 - \frac{\psi_A}{\psi_B} < \overline{\phi}_B$.

This ensures that $B$ (who we shall see holds the keys to separation) actually has a choice of inducing pooling or separation.

**Summary**

The exogenous parameters are $p$, $\underline{\phi}_A$, $\underline{\phi}_B$, $\overline{\phi}_A$, $\overline{\phi}_B$, $\psi_A$, and $\psi_B$. The endogenous choices are $\phi_A$, $\phi_B$, $m$, $s_A$, $s_B$, and $x$. The random variable is $\sigma$. As a sequential game of imperfect information, the natural equilibrium concept is perfect Bayesian equilibrium (PBE). I focus exclusively on pure-strategy PBE.

## Analysis

In Stage 2, from an analysis that is analogous to that in the baseline model, we have

$$\begin{aligned} s_A^*(\mu;\phi_A) &= \max\big\{\big(-(1-\phi_A)+\mu(2-\phi_A)\big)\psi_A, 0\big\}, \\ s_B^*(\mu;\phi_B) &= \max\big\{\big(1-\mu(2-\phi_B)\big)\psi_B, 0\big\}. \end{aligned}$$

Then the conditions required by a separating equilibrium are as follows:

$$(1-\phi_B)\psi_B \leq \psi_A, \tag{1}$$

$$(1-\phi_A)\psi_A \leq \psi_B. \tag{2}$$

Because $\psi_A \leq \psi_B$, it is immediate that Condition 2 is always satisfied. That is to say, $A$'s choice of $\phi_A$ never determines whether the separating equilibrium is possible. We therefore see that it is always a weakly dominant strategy for $A$ to select $\phi_A$ as large as possible. Whether we are in the separating or pooling equilibrium is in $B$'s hands, with separation occurring whenever $\phi_B$ is selected to satisfy Condition 1.[2] Analogous to $A$'s choice, then, selecting $\phi_B = 1 - \frac{\psi_A}{\psi_B}$ weakly dominates any $\phi_B < 1 - \frac{\psi_A}{\psi_B}$. That is to say, if pooling is going to happen, better that $\phi_B$ be as large as possible. This is summarized in the following

2. To ensure that an equilibrium exists, I assume that on the boundary at which the separating equilibrium comes into existence, the pooling equilibrium is still played.

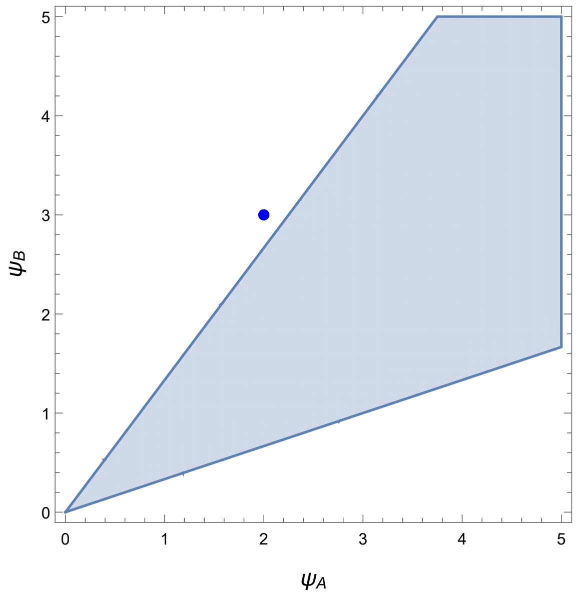



Figure B.1: An example in which $\psi_A = 2$, $\psi_B = 3$, $\phi_A = 1/4$, $\phi_B = 2/3$, and pooling occurs. Because $B$ can move the upper boundary of the cone, $\psi_B > \psi_A$ implies that $B$ is in control of whether separation is possible.

lemma:

**Lemma B.1** (Player strategies)**.** *It is a weakly dominant strategy for $A$ to set $\phi_A = \overline{\phi}_A$. For $B$, setting $\phi_B = 1 - \frac{\psi_A}{\psi_B}$ weakly dominates setting $\phi_B$ smaller.*

However, $B$ also realizes that $\phi_B$ even larger may bring about separation, at which point the specific choice of $\phi_B$ otherwise does not matter. Therefore, in determining the equilibrium, I consider $B$'s two candidates for optimal play. First, $B$ can select the largest $\phi_B$ that is still compatible with pooling. Second, $B$ can select anything larger than that to induce the separating equilibrium. Define

$$T_p' \equiv \frac{\psi_A\big(1-(2-\overline{\phi}_A)\overline{\phi}_A\big)+\psi_B}{\psi_A(2-\overline{\phi}_A)^2}.$$

We are now ready for the main result of this analysis:

**Proposition B.1** (Equilibrium outcomes)**.** *When $p \leq T'_p$, there exists a PBE in which $A$ sets $\phi_A = \overline{\phi}_A$, $B$ sets $\phi_B = \overline{\phi}_B$, and separation occurs. When $p \geq T'_p$, there exists a PBE in which $A$ sets $\phi_A = \overline{\phi}_A$, $B$ sets $\phi_B = 1 - \frac{\psi_A}{\psi_B}$, pooling occurs, and $s^*_B = 0$.*

A small value of $p$, then, means that $B$ prefers separation. That is, when $P$ is not overwhelmingly likely to be aligned with $A$, it benefits $B$'s policy goals more for both players to be able to identify their friends and enemies. And in keeping with $B$ having more capacity than $A$, notice that $T'_p \geq 1/2$, so even if $P$ is somewhat more likely to be aligned with $A$, it may still benefit $B$ to separate. When $p$ is large, it is remarkable that $B$ can induce pooling by setting $\phi_B$ sufficiently small but then does not end up having to grant any support at all. The mere presence of its superior, nonspecific resources proves tempting enough to opposition politicians such as to destroy any possibility for a separating equilibrium, thus preventing $A$ from being able to identify its friends and enemies.

I now look at comparative statics on $T'_p$. An increase in $T'_p$ means separation becomes more desirable for $B$, while a decrease means that pooling becomes more desirable:

**Proposition B.2** (Comparative statics)**.** *$T'_p$ increases in $\overline{\phi}_A$ and $\psi_B$ and decreases in $\psi_A$.*

Intuitively, as $B$'s capacity increases more relative to $A$, separation comes to benefit $B$ more. Finally, as $\overline{\phi}_A$ increases, $A$ is able to do increasingly well under pooling, eventually inducing $B$ to want to bring about separation.

# C Politician selection of complementarity extension

I extend the baseline model to examine how the politician might endogenously choose complementarity $\phi$. Selection of $\phi$ occurs before the baseline model plays out.

## Formal definition

Preliminaries are as in the baseline model. The sequence of moves is as follows:

**Stage 1**

1. $P$'s type $\sigma \in \{-1, 1\}$ is drawn and revealed to $P$. With probability $p \in (0, 1)$, $\sigma = -1$ and $P$ agrees with $A$. Otherwise, $P$ agrees with $B$.

2. $P$ selects $\phi \in [\underline{\phi}, \overline{\phi}]$.

Moves 2-5 from the baseline model shall comprise Stage 2. Utility functions are as in the baseline model. Assumption 1 is maintained. Next, I assume the following:

**Assumption C.1.** $\underline{\phi} \leq 1 - \frac{\psi_A}{\psi_B} < \overline{\phi}$.

This ensures that $P$ has a choice of values of $\phi$ that, given pooling in Stage 1, may correspond either to pooling or separation in Stage 2.

### Summary

The exogenous parameters are $p$, $\underline{\phi}$, $\overline{\phi}$, $\psi_A$, and $\psi_B$. The endogenous choices are $\phi$, $m$, $s_A$, $s_B$, and $x$. The random variable is $\sigma$. As a sequential game of imperfect information, the natural equilibrium concept is perfect Bayesian equilibrium (PBE). I focus exclusively on pure-strategy PBE. I further apply the *never dissuaded once convinced* refinement (Osborne and Rubenstein 1990, 96-8): once a group assigns probability one to any type, it does not engage in any further updating regardless of $P$'s subsequent actions.[3]

## Analysis

Analysis of Stage 2 is analogous to that in the baseline model. In the overall game, pooling may only occur if both types of $P$ would select the same value of $\phi$ and that value implies pooling in the baseline model.

Recalling Assumption 1, by Proposition 1 a value of $\phi$ implies pooling in the baseline model if and only if $(1-\phi)\psi_B \leq \psi_A$. This can be rearranged as $\phi \leq 1-\frac{\psi_A}{\psi_B}$. Considering this along with expressions for optimal support $s_A^*$ and $s_B^*$ in the baseline model, the following cases yield (presently setting aside the possibility of separation in Stage 1):

$$p \leq {}^1\!/_2 : \begin{cases} \phi \leq \min\left\{1-\frac{p}{1-p}, 1-\frac{\psi_A}{\psi_B}\right\} & \text{Pooling; } 0 = s_A^* < s_B^* \\ 1-\frac{p}{1-p} < \phi \leq 1-\frac{\psi_A}{\psi_B} & \text{Pooling; } 0 < s_A^*, s_B^* \\ 1-\frac{\psi_A}{\psi_B} < \phi & \text{Separation} \end{cases},$$

3. In the present setting, this appears to be a more reasonable refinement than that of Vincent (1998). Suppose instead that groups continue to update after the selection of $\phi$. Consider the case in which the prior probability of a politician aligned with group $I$ is small, and the capacity of $I$ is low. If the politician aligned with $I$ does appear, she would be able to select a small value of $\phi$ to ensure that she receives support from group $J$ under a subsequent pooling equilibrium, which may exceed the support that she would receive from $I$ under a subsequent separating equilibrium (this does not contradict Lemma 1, which relied on $\phi$ being fixed). But then it would have been sensible for both groups to continue to rely on the politician's selection of $\phi$ small to infer her type, given that only the politician aligned with $I$ has an incentive to do so.

$$p \geq 1/2 : \begin{cases} \phi \leq \min\left\{2 - \frac{1}{p}, 1 - \frac{\psi_A}{\psi_B}\right\} & \text{Pooling; } 0 = s_B^* < s_A^* \\ 2 - \frac{1}{p} < \phi \leq 1 - \frac{\psi_A}{\psi_B} & \text{Pooling; } 0 < s_A^*, s_B^* \\ 1 - \frac{\psi_A}{\psi_B} < \phi & \text{Separation} \end{cases}.$$

I reach the following result:

**Proposition C.1.** *One type of $P$ strictly prefers to select a value of $\phi$ that implies separation in Stage 2, such that separation in the overall game is guaranteed.*

# D Formal proofs

*Proof of Proposition 1.* Denote $P$ of type $\sigma = -1$ as $P_L$ and $P$ of type $\sigma = 1$ as $P_R$. A separating equilibrium takes the following form:

- Strategy for $P_L$: set $m = L$.
- Strategy for $P_R$: set $m = R$.
- Strategy for $I \in \{A, B\}$: grant $s_I^*(1)$ if $m = L$ and grant $s_I^*(0)$ otherwise.
- Beliefs: $\mu_L = 1$ and $\mu_R = 0$.

Holding fixed the behavior of groups, I check when both politician types have no incentive to deviate. The utility to $P_L$ from setting $m = L$ is $\psi_A$ while the utility to $P_L$ from misrepresenting and setting $m = R$ is $(1-\phi)\psi_B$. Then the utility of being truthful exceeds that of misrepresenting when $\frac{\psi_A}{\psi_B} \geq 1 - \phi$. Next, the utility to $P_R$ from setting $m = R$ is $\psi_B$ while the utility to $P_R$ from misrepresenting and setting $m = L$ is $(1-\phi)\psi_A$. Then the utility of being truthful exceeds that of misrepresenting when $\frac{\psi_B}{\psi_A} \geq 1 - \phi$ or equivalently, $\frac{\psi_A}{\psi_B} \leq \frac{1}{1-\phi}$. Taken together, this is $1 - \phi \leq \frac{\psi_A}{\psi_B} \leq \frac{1}{1-\phi}$. Given this, beliefs are consistent. Finally, $s_I^* : I \in \{A, B\}$ was already constructed to be optimal. $\square$

*Proof of Lemma 1.* Denote $P$ of type $\sigma = -1$ as $P_L$ and $P$ of type $\sigma = 1$ as $P_R$. Let superscript $\mathbb{S}$ denote separation and superscript $\mathbb{P}$ denote pooling. Expected utilities from separation are $\mathbb{E}U_{P_L}^{\mathbb{S}} = \psi_A$ and $\mathbb{E}U_{P_R}^{\mathbb{S}} = \psi_B$. Expected utilities from pooling are

$$\begin{aligned}
\mathbb{E}U_{P_L}^{\mathbb{P}} &= s_A^*(p) + (1-\phi)s_B^*(p) \\
&= \max\big\{\big(-(1-\phi) + p(2-\phi)\big)\psi_A, 0\big\} + (1-\phi)\big\{\big(1 - p(2-\phi)\big)\psi_B, 0\big\}, \\
\mathbb{E}U_{P_R}^{\mathbb{P}} &= (1-\phi)s_A^*(p) + s_B^*(p) \\
&= (1-\phi)\max\big\{\big(-(1-\phi) + p(2-\phi)\big)\psi_A, 0\big\} + \max\big\{\big(1 - p(2-\phi)\big)\psi_B, 0\big\}.
\end{aligned}$$

Recall the initial assumptions that $\psi_A \leq \psi_B$, $0 \leq t < 1$, and $0 < p < 1$. Next, by Proposition 1 and the hypothesis that separation is possible, it follows that $\psi_B \leq \frac{1}{1-\phi}\psi_A$. There are six possible cases: the Cartesian product of types of $P$ with contribution behavior under pooling ($p \leq \frac{1-\phi}{2-\phi}$ and only receiver $B$ contributes, $\frac{1-\phi}{2-\phi} < p < \frac{1}{2-\phi}$ and both receivers contribute, and $\frac{1}{2-\phi} \leq p$ and only receiver $A$ contributes). In each case, application of the assumptions along with $\psi_B \leq \frac{1}{1-\phi}\psi_A$ implies that $\mathbb{E}U_{P_I}^{\mathbb{S}} > \mathbb{E}U_{P_I}^{\mathbb{P}}$, with $I$ the corresponding type of $P$. □

*Proof of Proposition 2.* Follows from Proposition 4 of Farrell and Gibbons (1989) taken together with the present Lemma 1 (which substitutes for their Proposition 2, allowing application of the logic of Proposition 4 to the present case of continuous actions).

*Proof of Proposition 3.* Let $\mathbb{E}^{\tilde{\mathbb{S}}}[x|\psi_A, \psi_B]$ denote expected policy under a perfect information baseline and $\mathbb{E}^{\tilde{\mathbb{P}}}[x|\psi_A, \psi_B]$ denote expected policy when $P$ is banned from communicating. Notice of course that $\mathbb{E}^{\tilde{\mathbb{P}}}[x|\psi_A, \psi_B] = \mathbb{E}^{\mathbb{P}}[x|\psi_A, \psi_B]$, and $\mathbb{E}^{\tilde{\mathbb{S}}}[x|\psi_A, \psi_B] = \mathbb{E}^{\mathbb{S}}[x|\psi_A, \psi_B]$ when the separating equilibrium is supportable.

Expected policy under the perfect information benchmark is as follows:

$$\mathbb{E}^{\tilde{\mathbb{S}}}[x|\psi_A, \psi_B] = p(-\psi_A) + (1-p)\psi_B.$$

Expected policy under the no-communication benchmark is as follows:

$$\mathbb{E}^{\tilde{\mathbb{P}}}[x|\psi_A, \psi_B] = \begin{cases} (1-p(2-\phi))^2\psi_B & p \leq \frac{1-\phi}{2-\phi} \\ -(1-p(2-\phi)-\phi)^2\psi_A + (1-p(2-\phi))^2\psi_B & \frac{1-\phi}{2-\phi} \leq p \leq \frac{1}{2-\phi} \\ -(1-p(2-\phi)-\phi)^2\psi_A & \frac{1}{2-\phi} \leq p \end{cases}.$$

First observe that $\frac{\partial}{\partial \psi_A}\mathbb{E}^{\tilde{\mathbb{S}}}[x|\psi_A,\psi_B] = -p$. Next,

$$\frac{\partial}{\partial \psi_A}\mathbb{E}^{\tilde{\mathbb{P}}}[x|\psi_A,\psi_B] = \begin{cases} 0 & p \leq \frac{1-\phi}{2-\phi} \\ -(1-p(2-\phi)-\phi)^2 & \text{o/w} \end{cases},$$

so the proposition follows. □

*Proof of Proposition 4.* We must ask the conditions under which $\mathbb{E}^{\tilde{\mathbb{S}}}[x|(1-\phi)\psi_B,\psi_B] > \mathbb{E}^{\tilde{\mathbb{P}}}[x|(1-\phi)\psi_B,\psi_B]$. There are of course three cases: $p < \frac{1-\phi}{2-\phi}$, $\frac{1-\phi}{2-\phi} \leq p < \frac{1}{2-\phi}$, and $\frac{1}{2-\phi} \leq p$. In Cases 1 and 2, reduction of the system of inequalities consisting of the initial hypothesis and case (and basic initial assumptions of the model) demonstrates that the former always holds. In Case 3, the same process demonstrates that the initial hypothesis holds if and only if $p < \frac{1-\phi(1-\phi)}{2-\phi(3-\phi)}$. Because Cases 1 and 2 always imply $p < \frac{1-\phi(1-\phi)}{2-\phi(3-\phi)}$, the proposition follows. □

*Proof of Proposition 5.* The condition

$$\lim_{\psi_A \uparrow (1-\phi)\psi_B} \mathbb{E}[x] < \lim_{\psi_A \uparrow \frac{1}{1-\phi}\psi_B} \mathbb{E}[x]$$

is equivalent to

$$\mathbb{E}^{\tilde{\mathbb{P}}}[x|(1-\phi)\psi_B,\psi_B] < \mathbb{E}^{\tilde{\mathbb{S}}}\left[x\middle|\frac{1}{1-\phi}\psi_B,\psi_B\right].$$

There are three cases to consider: $p < \frac{1-\phi}{2-\phi}$, $\frac{1-\phi}{2-\phi} \leq p < \frac{1}{2-\phi}$, and $\frac{1}{2-\phi} \leq p$. In Cases 2 and 3, reduction of the system of inequalities consisting of the initial hypothesis and case (and basic initial assumptions of the model) demonstrates that the former never holds. In Case 1, the same process demonstrates that the initial hypothesis holds if and only if $p < \frac{1-2\phi}{2-\phi(3-\phi)}$. This is precisely the definition of separation strongly favoring $B$. Because each step in the chain of logical relationships was biconditional, the proposition follows. $\square$

*Proof of Lemma A.1.* Notice first that within pooling or separation, only the largest $\psi_B$ compatible with said equilibrium can be optimal.

In any Case, if $B$ cannot induce separation (i.e. $\psi_A < (1-\phi)\underline{\psi}_B$), it is clear that setting $\psi_B = \overline{\psi}_B$ is optimal. Suppose instead that $\psi_A \geq (1-\phi)\underline{\psi}_B$. Then $B$'s expected utility from separation (setting $\psi_B = \frac{1}{1-\phi}\psi_A$) is $\mathbb{E}U_B^{\mathbb{S}} = \frac{((1-p(2-\phi))\psi_A}{1-\phi}$.

Suppose that Case 1 holds. $B$'s expected utility from pooling (setting $\psi_B = \overline{\psi}_B$) is $\mathbb{E}U_B^{\mathbb{P}} = (1-p(2-\phi))^2\overline{\psi}_B$. Then $\mathbb{E}U_B^{\mathbb{S}} \geq \mathbb{E}U_B^{\mathbb{P}}$ implies (and is implied by) $\psi_A \geq (1-p(2-\phi))(1-\phi)\overline{\psi}_B$. Because $\psi_A < (1-\phi)\underline{\psi}_B$ makes separation infeasible for $B$ so that setting $\psi_B = \overline{\psi}_B$ must be optimal, it therefore follows that

$$\tilde{\psi}_A = \max\{(1-p(2-\phi))(1-\phi)\overline{\psi}_B, (1-\phi)\underline{\psi}_B\}.$$

Now suppose that Case 2 holds. $B$'s expected utility from pooling (setting $\psi_B = \overline{\psi}_B$) is

$$\mathbb{E}U_B^{\mathbb{P}} = -(1-p(2-\phi)-\phi)^2\psi_A + (1-p(2-\phi))^2\overline{\psi}_B.$$

Then $\mathbb{E}U_B^{\mathbb{S}} \geq \mathbb{E}U_B^{\mathbb{P}}$ implies (and is implied by)

$$\psi_A \geq \frac{(1-p(2-\phi))^2(1-\phi)}{(1-p)(2-\phi)(1-p(2-\phi)(1-\phi)-t(1-\phi))}\overline{\psi}_B, \tag{3}$$

so analogously to Case 1 it follows that

$$\tilde{\psi}_A = \max\left\{\frac{(1-p(2-\phi))^2(1-\phi)}{(1-p)(2-\phi)(1-p(2-\phi)(1-\phi)-t(1-\phi))}\overline{\psi}_B, (1-\phi)\underline{\psi}_B\right\}.$$

Suppose that Case 3 holds. $B$'s expected utility from pooling (setting $\psi_B = \overline{\psi}_B$) is

$$\mathbb{E}U_B^{\mathbb{P}} = -(1 - p(2 - \phi) - \phi)^2 \psi_A.$$

Then $\mathbb{E}U_B^{\mathbb{S}} \geq \mathbb{E}U_B^{\mathbb{P}}$ is equivalent to $p \leq \frac{1-\phi(1-\phi)}{2-\phi(3-\phi)}$, a condition unrelated to $\psi_A$. $\square$

*Proof of Proposition A.1.* Given what we know from Lemma A.1 about $B$'s choice of $\psi_B$, $A$'s expected utility from separation in any Case is $\mathbb{E}U_A^{\mathbb{S}} = \frac{(p(2-\phi)-1)\psi_A}{1-\phi}$. Then $\frac{d\mathbb{E}U_A^{\mathbb{S}}}{d\psi_A} = \frac{p(2-\phi)-1}{1-\phi}$, so it follows that $\frac{d\mathbb{E}U_A^{\mathbb{S}}}{d\psi_A} < 0$ in Cases 1 and 2, and $\frac{d\mathbb{E}U_A^{\mathbb{S}}}{d\psi_A} > 0$ in Case 3. Therefore, I conclude that if $A$ were to induce separation, in Cases 1 and 2, $A$ would set $\psi_A = \tilde{\psi}_A$ (if $B$ were ever so averse to separation such that $\tilde{\psi}_A > \overline{\psi}_A$, then $A$ simply cannot induce separation and sets $\psi_A = \overline{\psi}_A$). In Case 3, $A$ would set $\psi_A = \overline{\psi}_A$.

Note also that whenever $A$ desires pooling, $A$ sets $\psi_A$ as large as is compatible with this.

Suppose that Case 1 or 2 holds. Suppose first that $\tilde{\psi}_A \geq (1-\phi)\underline{\psi}_B$. Then at $\psi_A = \tilde{\psi}_A$, $A$ can induce either pooling or separation. But recall that $\tilde{\psi}_A$ is defined as the value of $\psi_A$ such that $B$ is indifferent between pooling and separation, and because the game in Stage 1 is constant-sum, this implies $A$'s indifference between pooling and separation (of course $B$ would not prefer separation with $\psi_A$ even greater). I conclude that $A$ sets $\psi_A = \tilde{\psi}_A$ and can assume that when indifferent, $A$ induces pooling.[4] Suppose instead that $\tilde{\psi}_A < (1-\phi)\underline{\psi}_B$. Because it was just demonstrated that $A$'s Stage 1 utility under separation is strictly decreasing in $\psi_A$, this implies that, since at $\tilde{\psi}_A$ $A$ is indifferent between pooling and separation, at $(1-\phi)\underline{\psi}_B$ $A$ must strictly prefer pooling. Then $A$ sets $\psi_A = \tilde{\psi}_A$ and induces pooling.

Suppose that Case 3 holds. Suppose that $B$ prefers pooling, i.e. $p \geq \frac{1-\phi(1-\phi)}{2-\phi(3-\phi)}$. Then $B$ always sets $\psi_B = \overline{\psi}_B$ regardless of $\psi_A$, so $A$ sets $\psi_A = \overline{\psi}_A$. Suppose instead that $B$

4. A lexicographic preference relation for $A$ by which $A$ first maximizes what is presently given as its Stage 1 utility function and next minimizes its Stage 2 cost of granting support would yield this as the optimum.

always prefers separation, i.e. $p \leq \frac{1-\phi(1-\phi)}{2-\phi(3-\phi)}$. Then $A$ can either induce pooling by setting $\psi_A = (1-\phi)\underline{\psi}_B$ or induce separation by setting $\psi_A = \overline{\psi}_A$. $A$'s utility from pooling is

$$\mathbb{E}U_A^{\mathbb{P}}\big((1-\phi)\underline{\psi}_B\big) = (1-p(2-\phi)-\phi)^2(1-\phi)\underline{\psi}_B,$$

while its utility from separation is $\mathbb{E}U_A^{\mathbb{S}}\big(\overline{\psi}_A\big) = \frac{\overline{\psi}_A(p(2-\phi)-1)}{1-\phi}$. Then $\mathbb{E}U_A^{\mathbb{S}} \geq \mathbb{E}U_A^{\mathbb{P}}$ implies (and is implied by)

$$\overline{\psi}_A \geq \frac{(1-p(2-\phi)-\phi)^2(1-\phi)^2}{p(2-\phi)-1}\underline{\psi}_B. \tag{4}$$

Then clearly $A$ induces separation by setting $\psi_A = \overline{\psi}_A$ if this condition holds and induces pooling by setting $\psi_A = (1-\phi)\underline{\psi}_B$ otherwise. Recalling that we are in Case 3 and $B$ always prefers separation, the condition can be rearranged as

$$p \geq \frac{\overline{\psi}_A - \sqrt{t\overline{\psi}_A\left(\overline{\psi}_A - 4(1-\phi)^2\underline{\psi}_B\right)}}{2(1-\phi)^2(2-\phi)\underline{\psi}_B} + \frac{1-\phi}{2-\phi}\ (= T_p). \tag{5}$$

Examining the right-hand side of Condition 4, observe that whenever $\phi > 0$, it follows that

$$\lim_{p \downarrow \frac{1}{2-\phi}} \frac{(1-p(2-\phi)-\phi)^2(1-\phi)^2}{p(2-\phi)-1}\underline{\psi}_B = \infty.$$

implying that approaching the boundary of Case 3 from within the case, Condition 4 is never satisfied. Next, if $\phi = 0$, to be in Case 3 we must have $p \geq 1/2$. Given this, $B$ is indifferent to separation rather than strictly dispreferring it (implying that $A$ is indifferent) only when $p = 1/2$. These observations imply that the right-hand side of Condition 5 must be greater than or equal to $\frac{1}{2-\phi}$. The proposition follows. □

*Proof of Proposition A.2.* The Condition 4 LHS increases in $\overline{\psi}_A$ and RHS increases in $\underline{\psi}_B$. □

*Proof of Lemma B.1.* As discussed, $A$'s choice of $\phi_A$ cannot determine whether pooling or separation occurs. If pooling occurs, $A$'s Stage 1 expected utility is

$$\mathbb{E}U_A^{\mathbb{P}} = \left((1-p)\phi_A - (1-2p)\right) s_A^*(p;\phi_A) - \left(p\phi_B - (2p-1)\right) s_B^*(p;\phi_B).$$

Suppose $p \leq 1/2$ and $\phi_A < \frac{1-2p}{1+p}$. Then $\frac{\partial \mathbb{E}U_A^{\mathbb{P}}}{\partial \phi_A} = 0$. Suppose instead that either $\phi_A > \frac{1-2p}{1+p}$ or $p \geq 1/2$ (or both). We have $\frac{\partial \mathbb{E}U_A^{\mathbb{P}}}{\partial \phi_A} = 2(1-p)((1-p)\phi_A - (1-2p))\psi_A > 0$. Then given that pooling occurs, $\phi_A = \overline{\phi}_A$ is always optimal. Given that separation occurs, $A$'s expected utility is not a function of $\phi_A$ and similarly, $\phi_A = \overline{\phi}_A$ is always optimal. A symmetric argument applies to $B$, except any $\phi_B > 1 - \frac{\psi_A}{\psi_B}$ leads to separation in Stage 2. □

*Proof of Proposition B.1.* Analysis of the Stage 2 subgame is as before. Next, Lemma B.1 tells us 1. $\phi_A = \overline{\phi}_A$ is always optimal for $A$ and 2. given that $B$ chooses to induce pooling, the largest such value of $\phi_B$ is selected, namely $1 - \frac{\psi_A}{\psi_B}$. We are left to determine which of two candidates is optimal for $B$: pooling with $\phi_B = 1 - \frac{\psi_A}{\psi_B}$ or separation with $\phi_B = \overline{\phi}_B$.

Utility to $B$ from separation is $\mathbb{E}U_B^{\mathbb{S}} = -p\psi_A + (1-p)\psi_B$. To determine utility to $B$ from pooling, allow two cases: $p \leq 1/2$ and $p > 1/2$. Suppose first that $p \leq 1/2$. Then utility from pooling is

$$\mathbb{E}U_B^{\mathbb{P}} = \frac{(-p\psi_A + (1-p)\psi_B)^2}{\psi_B} - c_A^*(p;\overline{\phi}_A)\big((1-p)\overline{\phi}_A - (1-2p)\big).$$

Given the assumed constraints on possible parameter values, $\mathbb{E}U_B^{\mathbb{S}} \geq \mathbb{E}U_B^{\mathbb{P}}$ must follow.

Suppose instead that $p > 1/2$. Then utility from pooling is

$$\mathbb{E}U_B^{\mathbb{P}} = \frac{(-p\psi_A + (1-p)\psi_B)c_B^*\big(p; 1 - \frac{\psi_A}{\psi_B}\big)}{\psi_B} - \psi_A\big((1-p)\overline{\phi}_A - (1-2p)\big)^2.$$

Then $\mathbb{E}U_B^{\mathbb{S}} \geq \mathbb{E}U_B^{\mathbb{P}}$ implies (and is implied by) $p \leq T_p'$.

Suppose that $p \geq T_p'$ and $B$ induces pooling. To see that $c_B^* = 0$, observe that $c_B^*\big(p; 1 - \frac{\psi_A}{\psi_B}\big) > 0$ implies $p < \frac{\psi_B}{\psi_A + \psi_B}$, which contradicts $p \geq T_p'$.

Finally, observing that $T_p' > 1/2$, I find that $T_p'$ is always the threshold dividing the region of $p$ in which the specified separating equilibrium exists from that in which the specified pooling equilibrium exists. □

*Proof of Proposition B.2.* We have $\frac{\partial T_p'}{\partial \psi_A} = -\frac{\psi_B}{\psi_A^2(2-\overline{\phi}_A)^2} < 0$, $\frac{\partial T_p'}{\partial \psi_B} = \frac{1}{\psi_A(2-\overline{\phi}_A)^2} > 0$, and $\frac{\partial T_p'}{\partial \overline{\phi}_A} = \frac{2(\psi_B - (1-\overline{\phi}_A)\psi_A)}{\psi_A(\overline{\phi}_A - 2)^3} > 0$. □

*Proof of Proposition C.1.* **Case 1: $p \leq 1/2$.** Suppose that $P_B$ selects $\phi \leq \min\left\{1 - \frac{p}{1-p}, 1 - \frac{\psi_A}{\psi_B}\right\}$. Expected utility to $P_B$ is

$$\mathbb{E}U_{P_B} = \big(\phi p + (1-2p)\big)\psi_B,$$

which is maximized at $\phi = \min\left\{1 - \frac{p}{1-p}, 1 - \frac{\psi_A}{\psi_B}\right\}$. Suppose next that $P_B$ selects $\phi \in \left(1 - \frac{p}{1-p}, 1 - \frac{\psi_A}{\psi_B}\right]$. Expected utility to $P_B$ is

$$\mathbb{E}U_{P_B} = (1-\phi)\big((1-p)\phi + 2p - 1\big)\psi_A + \big(\phi p + (1-2p)\big)\psi_B. \tag{6}$$

This has a critical point at

$$\phi^* = \frac{(2-3p)\psi_A + p\psi_B}{2(1-p)\psi_A}. \tag{7}$$

The second derivative test demonstrates that this is globally concave in $\phi$. But notice that $\phi^* > 1 - \frac{\psi_A}{\psi_B}$, such that if any value of $\phi \in \left[0, 1 - \frac{\psi_A}{\psi_B}\right]$ were optimal, it must be $1 - \frac{\psi_A}{\psi_B}$. Then we are left to compare expected utility from pooling at $1 - \frac{\psi_A}{\psi_B}$ to that from separation. The former is Expression 6 setting $\phi = 1 - \frac{\psi_A}{\psi_B}$. The latter is simply $\psi_B$. The latter is strictly greater than the former, such that separation is always strictly preferred.

**Case 2:** $p \geq 1/2$**.** Suppose that $P_B$ selects $\phi \leq \min\left\{2 - \frac{1}{p}, 1 - \frac{\psi_A}{\psi_B}\right\}$. Expected utility to $P_B$ is

$$\mathbb{E}U_{P_B} = (1-\phi)\big((1-p)\phi + 2p - 1\big)\psi_A.$$

This has a critical point at

$$\phi^* = \frac{2-3p}{2(1-p)}.$$

The second derivative test demonstrates that this is globally concave in $\phi$, but note for later that $\phi^* < 2 - \frac{1}{p}$ coincides with $p > 2 - \sqrt{2}$.

Suppose next that $P_B$ selects $\phi \in \left(2 - \frac{1}{p}, 1 - \frac{\psi_A}{\psi_B}\right]$. Expected utility to $P_B$ is as in Expression 6, and so the critical point is as in Expression 7. But as before, $\phi^* > 1 - \frac{\psi_A}{\psi_B}$.

We are left to perform three expected utility comparisons: separation vs. 1. pooling at $\phi = \frac{2-3p}{2(1-p)}$ (when $p > 2 - \sqrt{2}$), 2. pooling at $\phi = 2 - \frac{1}{p}$ (when $p \leq 2 - \sqrt{2}$), and 3. pooling at $1 - \frac{\psi_A}{\psi_B}$. The third comparison was already performed in Case 1, demonstrating that separation is strictly preferred. Performing the second comparison also shows that separation is strictly preferred. The first comparison implies that pooling is preferred if and only if $\psi_B \leq \frac{p^2}{4(1-p)}\psi_A$.

The final step, then, is to determine what $P_A$ prefers to do when $\psi_B \leq \frac{p^2}{4(1-p)}\psi_A$ and $p > 2 - \sqrt{2}$. Suppose that $P_A$ selects $\phi \leq \min\left\{2 - \frac{1}{p}, 1 - \frac{\psi_A}{\psi_B}\right\}$. Expected utility to $P_A$ is

$$\mathbb{E}U_{P_A} = \big((1-p)\phi + 2p - 1\big)\psi_A,$$

which is clearly increasing in $\phi$, implying a maximum at the upper corner. Suppose next that $P_A$ selects $\phi \in \left(2 - \frac{1}{p}, 1 - \frac{\psi_A}{\psi_B}\right]$. Expected utility to $P_A$ is

$$\mathbb{E}U_{P_A} = \big(p - (1-\phi)(1-p)\big)\psi_A + (1-\phi)\big(1 - p(2-\phi)\big)\psi_B. \tag{8}$$

This has a critical point at

$$\phi^* = \frac{(1-p)\psi_A + (3p-1)\psi_B}{2p\psi_B}. \tag{9}$$

The second derivative test demonstrates that this is globally concave in $\phi$. But notice that $\phi^* > 1 - \frac{\psi_A}{\psi_B}$, such that if any value of $\phi \in \left[0, 1 - \frac{\psi_A}{\psi_B}\right]$ were optimal, it must be $1 - \frac{\psi_A}{\psi_B}$. Then we are left to compare expected utility from pooling at $1 - \frac{\psi_A}{\psi_B}$ to that from separation. The former is Expression 8 setting $\phi = 1 - \frac{\psi_A}{\psi_B}$. The latter is simply $\psi_A$. The latter is strictly greater than the former, such that separation is always strictly preferred.

Then one type of $P$ always strictly prefers separation. This can always be induced by selecting $\phi$ sufficiently large, such that failure to do so is informative in itself. $\square$